Mathematical Modelling and Numerical Analysis Will be set by the publisher
Modélisation Mathématique et Analyse Numérique

AN ADAPTIVE PARAMETRIZED-BACKGROUND DATA-WEAK APPROACH TO VARIATIONAL DATA ASSIMILATION

Tommaso Taddei [1]

Abstract. We present an Adaptive Parametrized-Background Data-Weak (APBDW) approach to the variational data assimilation (state estimation) problem. The approach is based on the Tikhonov regularization of the PBDW formulation [Y Maday, AT Patera, JD Penn, M Yano, Int J Numer Meth Eng, 102(5), 933-965], and exploits the connection between PBDW and kernel methods for regression. An adaptive procedure is presented to handle the experimental noise. *A priori* and *a posteriori* estimates for the L^2 state-estimation error motivate the approach and guide the adaptive procedure. We present results for two synthetic model problems to illustrate the elements of the methodology. We also consider an experimental thermal patch configuration to demonstrate the applicability of our approach to real physical systems.

Résumé. Nous présentons une approche Adaptive-Parameterized-Background Data-Weak (APBDW) pour le problème d'assimilation de données variationnelles. L'approche est fondée sur la régularisation Tychonoff de la formulation PBDW [Y Maday, AT Patera, JD Penn, M Yano, Int J Numer Meth Eng, 102(5), 933-965], et consiste en une procédure adaptative pour considérer le bruit expérimental. Des estimations a priori et a posteriori de l'état L^2 (estimation d'erreur) motivent l'approche et servent de guide à la procédure adaptative. Nous présentons des résultats pour deux problèmes de modèle synthétique pour illustrer les éléments de la méthodologie. Nous considérons aussi une configuration expérimentale de patch thermique pour montrer que notre approche est applicable dans le cadre de systèmes physiques.

1991 Mathematics Subject Classification. 62-07,93E24.

March 5, 2016.

1. Introduction

Data assimilation refers to the estimation of the state u^{true} of a physical system over the domain of interest $\Omega \subset \mathbb{R}^d$ by combining experimental data with a mathematical model of the dynamics of the system. For real-time and in situ applications, data assimilation techniques should provide an estimate of the state rapidly with little or no communication with extensive offline resources. Furthermore, for safety reasons, it is key to certify the reliability of our estimate using either probabilistic (i.e., confidence intervals) or deterministic (i.e., error bounds) approaches.

Keywords and phrases: variational data assimilation; parametrized partial differential equations; model order reduction; kernel methods

[1] Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA, ttaddei@mit.edu

The goal of this work is to develop a variational data assimilation procedure that combines a parameterized best-knowledge mathematical model and experimental data to rapidly obtain a reliable estimate of the state $u^{true} \in \mathcal{U}$ over Ω. We denote by $\{y_m\}_{m=1}^M$ the set of experimental measurements, and we denote by $u^{bk}(\mu) \in \mathcal{U}$ the solution to our parameterized best-knowledge mathematical model for the parameter value $\mu \in \mathcal{P}$. Here, the space $\mathcal{U} = \mathcal{U}(\Omega)$ is a suitable Hilbert space defined over Ω, while $\mathcal{P} \subset \mathbb{R}^P$ reflects the uncertainty in the value of the parameters associated with the model. Since experimental apparatuses are typically affected by errors, measurements are in general of the form $y_m = \ell_m(u^{true}) + \epsilon_m$, where $\ell_m : \mathcal{U} \mapsto \mathbb{R}$ is a known functional and ϵ_m reflects the observational noise. On the other hand, the uncertainty in the parameters of the model leads to the definition of the best-knowledge manifold $\mathcal{M}^{bk} := \{u^{bk}(\mu) : \mu \in \mathcal{P}\} \subset \mathcal{U}$, which collects the solution to the parameterized best-knowledge model for all values of the parameter in $\mathcal{P}$.

In [24,25], Maday et al. introduced the so-called Parameterized-Background Data-Weak (PBDW) approach. The key idea of the PBDW formulation is to seek an approximation $u^\star = z^\star + \eta^\star$ to the true field u^{true} employing projection-by-data. The first contribution to $u^\star$, $z^\star \in \mathcal{Z}_N$, is the "deduced background estimate." The linear N-dimensional space $\mathcal{Z}_N \subset \mathcal{U}$ is informed by the best-knowledge manifold $\mathcal{M}^{bk}$, which we hope is close to the true field. The second contribution to $u^\star$, $\eta^\star \in \mathcal{U}_M$, is the "update estimate". The linear M-dimensional space $\mathcal{U}_M$ is the span of the Riesz representations of the M observation functionals $\{\ell_m\}_{m=1}^M$. While the background estimate incorporates our *a priori* knowledge of the state, the update addresses the deficiencies of the best-knowledge model by improving the approximation properties of the search space. Projection-by-data, as opposed to projection-by-model, implies that the parameterized model is not directly used during the data assimilation procedure. This feature significantly simplifies the computational procedure and guarantees real-time responses.

In this work, we present an adaptive Parameterized-Background Data-Weak (APBDW) approach that extends the original PBDW formulation to the case of pointwise noisy measurements; $\ell_m := \delta_{x_m}$ for some $x_m \in \Omega$, $m = 1, \ldots, M$. Our approach is based on the Tikhonov regularization of the PBDW formulation and relies on an adaptive procedure for the selection of the hyper-parameters. The adaptive procedure chooses the hyper-parameters that minimize an estimate of the L^2 state-estimation error on a validation dataset. The extension to pointwise measurements is based on the theory of Reproducing Kernel Hilbert Spaces (RKHS, [1]) and exploits the connection between the PBDW formulation and kernel methods for regression ([9, 30]) We also present *a priori* and *a posteriori* error estimates for measurements affected by either systematic or homoscedastic random error. These estimates motivate the approach from a theoretical standpoint, and guide the adaptive procedure.

Our approach shares some key features with a number of existing techniques from the statistical learning and data assimilation literature. In the statistical learning literature, our approach is equivalent to the approach presented in [16] by Kimeldorf and Wahba. However, in their paper, the authors did not relate the background space $\mathcal{Z}_N$ to the solution manifold associated with a parameterized PDE and they did not discuss how to practically build the space based on the available prior knowledge about the state. In the data assimilation literature, our work is related to the approach presented by Bennett in [2, 3], and to the so-called 3D-VAR, first introduced by Lorenc [21, 22] for steady problems, then extended to time dependent problems with the name of 4D-VAR ([8]) and more recently coupled with model order reduction techniques ([6,31]) to reduce the computational costs. Although the formulations proposed by Bennett and Lorenc are informed by the solution to a best-knowledge differential model, both approaches are not naturally designed to include a parameterized background space in the formulation. The *a priori* error analysis can be seen as a generalization of the work of Krebs et al. [18], while the *a posteriori* error analysis is related to [29]. Finally, the adaptive procedure employed in this work, which is known in statistics as holdout validation (see, e.g., [15, Chapter 7] and [17]), is strongly connected to the approach proposed in [25, section 5.8] to improve the approximation properties of the background space $\mathcal{Z}_N$ for sequential data assimilation.

The outline of the paper is as follows. In section 2, we present the formulation and the well-posedness analysis. We further relate our formulation to a number of other methods proposed in the data assimilation and statistical learning literature. Then, in section 3, we present *a priori* and *a posteriori* estimates for the $L^2(\Omega)$ state-estimation error. In section 4, we exploit the error analysis to design an adaptive procedure for the

selection of the parameters. In section 5, we present numerical results for two synthetic problems. Finally, in section 6, we present the results for a physical thermal patch problem.

2. Formulation

2.1. Preliminaries

By way of preliminaries, we introduce notation used throughout the paper. Given the Lipschitz domain $\Omega \subset \mathbb{R}^d$, we introduce the Reproducing Kernel Hilbert Space (RKHS) $\mathcal{U}$ endowed with the inner product $(\cdot,\cdot)$ and the induced norm $\|\cdot\| = \sqrt{(\cdot,\cdot)}$. We further denote by $K : \Omega\times\Omega \mapsto \mathbb{R}$ the reproducing kernel associated with $\mathcal{U}$, which is assumed to be known explicitly. Finally, for any closed subspace $\mathcal{Q} \subset \mathcal{U}$, we denote by $\Pi_{\mathcal{Q}} : \mathcal{U} \mapsto \mathcal{Q}$ the orthogonal projection operator onto $\mathcal{Q}$, and we denote by $\mathcal{Q}^{\perp}$ its orthogonal complement.

We now introduce the dataset of experimental observations considered in this work. We consider the dataset $\mathcal{D}_M := \{(x_m, y_m)\}_{m=1}^M$ where $\mathcal{X}_M := \{x_1, \dots, x_M\} \subset \Omega$ represents the set of observation points, while $\mathcal{Y}_M := \{y_1, \dots, y_M\} \subset \mathbb{R}$ are approximations of the true states in the observation points, $y_m \approx u^{true}(x_m)$.

We briefly provide the definition of RKHS and we recall a number of properties exploited in this work. For the purpose of this paper, a RKHS is an Hilbert space such that the evaluation functional δ_x belongs to the dual space of $\mathcal{U}$ for any $x \in \Omega$. If we denote by $\phi : x \in \Omega \mapsto \phi(x) \in \mathcal{U}$ the feature map associated with $\mathcal{U}$ such that $(\phi(x), f) = f(x)$ for any $x \in \Omega$, the Reproducing Kernel associated with $\mathcal{U}$ is given by $K(x, y) = (\phi(x), \phi(y))$. It is easy to verify that K is symmetric, $K(x, y) = K(y, x)$ for all $x, y \in \Omega$. It is also easy to verify that K is positive definite that is

$$\sum_{i,j=1}^{N} c_i\, c_j K(x_j, x_i) \; > 0$$

for all distinct $\{x_i\}_{i=1}^N \subset \Omega$, for all $\mathbf{c} \in \mathbb{R}^N \setminus \{\mathbf{0}\}$ and for any $N > 0$.

Recalling Moore-Aronszajn theorem ([1]), there exists a duality between symmetric positive definite kernels and RKHS: we have indeed that given a symmetric positive definite (SPD) kernel K there exists a RKHS for which K is the reproducing kernel. This RKHS is referred to as *native space* of K. We can thus first choose an *explicit* SPD kernel and then exploit Moore-Aronszajn theorem to recover the variational interpretation.

2.2. Adaptive PBDW statement

We introduce some definitions. First, we define the N-dimensional background space $\mathcal{Z}_N = \text{span}\{\zeta_n\}_{n=1}^N \subset \mathcal{U}$. The space $\mathcal{Z}_N$ encodes our prior knowledge about the state u^{true}; we refer to section 2.5.1 for practical strategies for the construction of the background space. Given the dataset $\mathcal{D}_M$, we further introduce the empirical risk $V_M : \mathcal{U} \mapsto \mathbb{R}$:

$$V_M(u) = \frac{1}{M} \sum_{m=1}^{M} (u(x_m) - y_m)^2. \tag{1}$$

We have now the elements to state the Adaptive Parameterized-Background Data-Weak (APBDW) formulation: given $\xi > 0$, find $u_\xi^\star \in \mathcal{U}$ such that

$$u_\xi^\star = \arg\min_{u\in\mathcal{U}} \; J_\xi(u) := \xi \|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + V_M(u). \tag{2}$$

To simplify notation, in the state estimate $u_\xi^\star$ we omit the subscripts associated with the discretization parameters M and N.

2.3. Well-posedness analysis

Proposition 2.1 provides a sufficient condition for the well-posedness of problem (2), and an explicit low-dimensional representation for the solution that allows real-time calculations. The result was first proved by Kimeldorf and Wahba in [16] in a slightly different form. Proposition 2.4 shows that problem (2) can be

reformulated as a two-field minimization statement. The latter result will serve to generalize our statement to the case in which the background does not belong to $\mathcal{U}$.

Proposition 2.1. *Let $\mathcal{U}_M := span\{K_{x_m}\}_{m=1}^M$ and let $\beta_{N,M}$ be defined as*

$$\beta_{N,M} := \inf_{z\in\mathcal{Z}_N} \sup_{v\in\mathcal{U}_M} \frac{(z,v)}{\|z\|\|v\|}. \tag{3}$$

Then, if $\beta_{N,M} > 0$, the solution $u_\xi^\star$ to (2) exists and is unique. Furthermore, the solution to (2) is of the form

$$u_\xi^\star(x) = \eta_\xi^\star + z_\xi^\star, \tag{4}$$

where $\eta_\xi^\star \in \mathcal{U}_M \cap \mathcal{Z}_N^\perp$ and $z_\xi^\star \in \mathcal{Z}_N$.

In view of the proof of Proposition 2.1, we state and prove two lemmas.

Lemma 2.2. *([36]) Let $\mathcal{U}_M := span\{K_{x_m}\}_{m=1}^M$ and let $\beta_{N,M}$ be defined as in (3). Then, we have that*

$$\beta_{N,M} = \inf_{\eta\in\mathcal{U}_M^\perp} \frac{\|\Pi_{\mathcal{Z}_N^\perp}\eta\|}{\|\eta\|}. \tag{5}$$

Proof. To simplify notation, given the linear space $\mathcal{Q}$, we define $\mathcal{Q}^{(1)} = \{q\in\mathcal{Q}:\ \|q\| = 1\}$. We now prove (5).

$$\begin{aligned}
\beta_{N,M}^2 = &\left(\inf_{z\in\mathcal{Z}_N^{(1)}} \sup_{v\in\mathcal{U}_M^{(1)}} (z,v)\right)^2 = \inf_{z\in\mathcal{Z}_N^{(1)}} \|\Pi_{\mathcal{U}_M} z\|^2 = 1 - \sup_{z\in\mathcal{Z}_N^{(1)}} \|\Pi_{\mathcal{U}_M^\perp} z\|^2 \\
= &\ 1 - \left(\sup_{z\in\mathcal{Z}_N^{(1)}} \sup_{q\in\mathcal{U}_M^{\perp(1)}} (z,q)\right)^2 = 1 - \left(\sup_{q\in\mathcal{U}_M^{\perp(1)}} \sup_{z\in\mathcal{Z}_N^{(1)}} (z,q)\right)^2 \\
= &\ 1 - \sup_{q\in\mathcal{U}_M^{\perp(1)}} \|\Pi_{\mathcal{Z}_N} q\|^2 = \inf_{q\in\mathcal{U}_M^{\perp(1)}} \|\Pi_{\mathcal{Z}_N^\perp} q\|^2
\end{aligned}$$

Thesis follows. □

Lemma 2.3. *Let $\mathcal{U}_{M'} := span\{K_{x_m}\}_{m=1}^{M'}$, $M' \le M$. Let us introduce $\beta_{N,M'} = \inf_{z\in\mathcal{Z}_N} \sup_{v\in\mathcal{U}_{M'}} \frac{(z,v)}{\|z\|\|v\|}$, and the matrix $\mathbb{K}^{(M')} \in \mathbb{R}^{M',M'}$, $\mathbb{K}_{m,m'}^{(M')} = K(x_m, x_{m'})$. Let us further define*

$$c_{N,M} := \max_{M'=1,\dots,M} \hat{c}_{N,M'}, \qquad \hat{c}_{N,M'} = \min\left(\frac{1}{2}\lambda_{min}(\mathbb{K}^{(M')}),\ \frac{\lambda_{min}(\mathbb{K}^{(M')})}{2+\lambda_{min}(\mathbb{K}^{(M')})}\,\beta_{N,M'}^2\right), \tag{6a}$$

where $\lambda_{min}(\mathbb{K}^{(M')})$ denotes the minimum eigenvalue of the matrix $\mathbb{K}^{(M')}$.

Then, the following bound holds:

$$\widetilde{J}(u) \ = \ \|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + \sum_{m=1}^M u(x_m)^2 \ge c_{N,M}\|u\|^2, \qquad \forall\, u\in\mathcal{U}. \tag{6b}$$

Proof. We first claim that for any M' such that $\beta_{N,M'} > 0$ we have

$$\widetilde{J}_{M'}(u) \ = \ \|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + \sum_{m=1}^{M'} u(x_m)^2 \ge \hat{c}_{N,M'}\|u\|^2, \qquad \forall\, u\in\mathcal{U}. \tag{7}$$

Given (7), we find that

$$\widetilde{J}(u) \ge \widetilde{J}_{M'}(u) \ge \hat{c}_{N,M'}\|u\|^2 \ \ \forall\, M' \le M \ \Rightarrow\ \widetilde{J}(u) \ge \left(\max_{M'} \hat{c}_{N,M'}\right)\|u\|^2,$$

which is the thesis.

We now show (7). Given $u \in \mathcal{U}$, we introduce $u_1 = \Pi_{\mathcal{U}_{M'}^{\perp}} u$, $u_2 = \Pi_{\mathcal{U}_{M'}} u = \sum_{m=1}^{M'} (\mathbf{u}_2)_m K_{x_m}$. Then, we observe that

$$u_1(x_m) = (\underbrace{K_{x_m}}_{\in \mathcal{U}_{M'}}, u_1) = 0, \qquad m = 1, \dots, M'. \tag{8}$$

We further observe that

$$\sum_{m=1}^{M'} u_2(x_m)^2 = \|\mathbb{K}^{(M')}\mathbf{u}_2\|_2^2, \qquad \|u_2\|^2 = \mathbf{u}_2^T\, \mathbb{K}^{(M')}\, \mathbf{u}_2,$$

which implies that

$$\min_{u_2 \in \mathcal{U}_{M'}} \frac{\sum_{m=1}^{M'} u_2(x_m)^2}{\|u_2\|^2} = \min_{\mathbf{u}_2 \in \mathbb{R}^{M'}} \frac{\|\mathbb{K}^{(M')}\mathbf{u}_2\|_2^2}{\mathbf{u}_2^T\, \mathbb{K}^{(M')}\, \mathbf{u}_2} = \lambda_{min}(\mathbb{K}^{(M')}). \tag{9}$$

Combining (8) and (9), we obtain

$$\sum_{m=1}^{M'} u(x_m)^2 = \sum_{m=1}^{M'} u_2(x_m)^2 \geq \lambda_{min}(\mathbb{K}^{(M')})\, \|u_2\|^2.$$

Now, recalling the identity $2ab \geq -\frac{1}{\epsilon}a^2 - \epsilon b^2$ valid for any $\epsilon > 0$, and Lemma 2.2, we obtain:

$$\begin{aligned} \widetilde{J}_{M'}(u) = \widetilde{J}_{M'}(u_1 + u_2) \geq & \quad \|\Pi_{\mathcal{Z}_N^{\perp}} u_1\|^2 + \|\Pi_{\mathcal{Z}_N^{\perp}} u_2\|^2 + 2\,(\Pi_{\mathcal{Z}_N^{\perp}} u_1, \Pi_{\mathcal{Z}_N^{\perp}} u_2) + \lambda_{min}(\mathbb{K}^{(M')})\, \|u_2\|^2 \\ \geq & \quad (1-\epsilon)\, \beta_{N,M'}^2 \|u_1\|^2 + \left(1 - \tfrac{1}{\epsilon}\right) \|\Pi_{\mathcal{Z}_N^{\perp}} u_2\|^2 + \lambda_{min}(\mathbb{K}^{(M')}) \|u_2\|^2 \end{aligned}$$

Let us consider $\epsilon \in \left(\frac{1}{1+\lambda_{min}(\mathbb{K}^{(M')})}, 1\right)$. Recalling that $\|\Pi_{\mathcal{Z}_N^{\perp}} u_2\| \leq \|u_2\|$, we obtain

$$\begin{aligned} \widetilde{J}_{M'}(u) \geq & \quad (1-\epsilon)\beta_{N,M'}^2\, \|u_1\|^2 + \left(\lambda_{min}(\mathbb{K}^{(M')}) + 1 - \tfrac{1}{\epsilon}\right) \|u_2\|^2 \\ \geq & \quad \min\left(\lambda_{min}(\mathbb{K}^{(M')}) + 1 - \tfrac{1}{\epsilon},\, (1-\epsilon)\beta_{N,M'}^2\right) \underbrace{(\|u_1\|^2 + \|u_2\|^2)}_{=\|u\|^2}. \end{aligned}$$

Estimate (7) follows by considering $\epsilon = \frac{2}{2+\lambda_{min}(\mathbb{K}^{(M')})}$. □

We observe that $c_{N,M}$ is monotonic increasing with M; therefore, it is asymptotically bounded from below in the limit $M \to \infty$.

Proof. (Proposition 2.1) Applying Lemma 2.3, we find that the objective function $J_\xi : \mathcal{U} \mapsto \mathbb{R}$ is strictly convex. Existence and uniqueness then follow from [10, Theorem 3, Chapter 8.2].

We now show the decomposition (4). If we define $\eta_\xi^\star = \Pi_{\mathcal{Z}_N^{\perp}} u_\xi^\star$ and $z_\xi^\star = \Pi_{\mathcal{Z}_N} u_\xi^\star$, this corresponds to show that $\eta_\xi^\star \in \mathcal{U}_M$. Recalling that $V_M(u) = V_M(\Pi_{\mathcal{U}_M} u)$ for all $u \in \mathcal{U}$, we find

$$J_\xi(u_\xi^\star) = J_\xi(\eta_\xi^\star + z_\xi^\star) \geq J_\xi(\Pi_{\mathcal{U}_M}\eta_\xi^\star + z_\xi^\star).$$

Thus, since $u_\xi^\star$ is the unique minimizer, we must have $\Pi_{\mathcal{U}_M}\eta_\xi^\star = \eta_\xi^\star$. Thesis follows. □

Proposition 2.4. *Suppose that $\mathcal{Z}_N \subset \mathcal{U}$ and $\beta_{N,M} > 0$. Then, $\eta_\xi^\star$ and $z_\xi^\star$ defined in* (4) *solve the following problem:*

$$\min_{(\eta,z) \in \mathcal{U} \times \mathcal{Z}_N} J_\xi^{(2)}(\eta, z) := \xi\|\eta\|^2 + V_M(\eta + z). \tag{10}$$

Proof. Let $(\eta, z) \in \mathcal{U} \times \mathcal{Z}_N$. Then, we have

$$\begin{aligned} J_\xi^{(2)}(\eta, z) = &\quad \xi \left(\|\Pi_{\mathcal{Z}_N^\perp} \eta\|^2 + \|\Pi_{\mathcal{Z}_N} \eta\|^2 \right) + V_M(\Pi_{\mathcal{Z}_N} \eta + \Pi_{\mathcal{Z}_N^\perp} \eta + z) \\ = &\quad \xi \|\Pi_{\mathcal{Z}_N} \eta\|^2 \, + \, J_\xi^{(2)}(\Pi_{\mathcal{Z}_N^\perp} \eta, \Pi_{\mathcal{Z}_N} \eta + z) \geq J_\xi^{(2)}(\Pi_{\mathcal{Z}_N^\perp} \eta, \Pi_{\mathcal{Z}_N} \eta + z). \end{aligned}$$

Therefore, we have that

$$\min_{(\eta,z) \in \mathcal{U} \times \mathcal{Z}_N} \; J_\xi^{(2)}(\eta, z) = \min_{(\eta,z) \in \mathcal{Z}_N^\perp \times \mathcal{Z}_N} \; J_\xi^{(2)}(\eta, z)$$

Thesis follows by observing that $J_\xi(u) = J_\xi^{(2)}(\Pi_{\mathcal{Z}_N^\perp} u, \Pi_{\mathcal{Z}_N} u)$. □

We observe that the two-field functional $J_\xi^{(2)}(\eta, z)$ is well-defined for $\mathcal{Z}_N \subset C(\Omega)$. We thus expect that it is possible to derive sufficient conditions for the well-posedness of (10) without assuming that $\mathcal{Z}_N \subset \mathcal{U}$. We address this issue in section 2.4.

2.4. **Algebraic formulation**

We now discuss how to solve problem (2) numerically. With this in mind, we first introduce the matrices $\mathbb{K} \in \mathbb{R}^{M,M}$, $\mathbb{Z} \in \mathbb{R}^{N,N}$, $\mathbb{L} \in \mathbb{R}^{M,N}$ such that

$$\mathbb{K}_{m,m'} = (K_{x_m}, K_{x_{m'}}) = K(x_m, x_{m'}), \;\; \mathbb{Z}_{n,n'} = (\zeta_n, \zeta_{n'}), \;\; \mathbb{L}_{m,n} = (K_{x_m}, \zeta_n) = \zeta_n(x_m). \tag{11}$$

Proposition 2.5 provides the algebraic formulation of the APBDW statement (2).

Proposition 2.5. *Assuming $\beta_{N,M} > 0$, the solution $u_\xi^\star$ to (2) is given by*

$$u_\xi^\star(x) = \sum_{m=1}^{M} \eta_m^\star K_{x_m}(x) \, + \, \sum_{n=1}^{N} z_n^\star \zeta_n(x). \tag{12}$$

the pair $(\boldsymbol{\eta}^\star, \boldsymbol{z}^\star) \in \mathbb{R}^M \times \mathbb{R}^N$ solves

$$\begin{cases} (\xi M \mathbb{I} + \mathbb{K}) \, \boldsymbol{\eta}^\star + \mathbb{L} \mathbf{z}^\star &= \mathbf{y} \\ \mathbb{L}^T \, \boldsymbol{\eta}^\star &= \mathbf{0}. \end{cases} \tag{13}$$

In anticipation of the proof, we recall a standard result (see, e.g., [26, Section 1.3.5]).

Lemma 2.6. *The inf-sup constant $\beta_{N,M}$ is the square root of the minimum eigenvalue of the following eigenproblem:*

$$\mathbb{L}^T \, \mathbb{K}^{-1} \, \mathbb{L} \, \mathbf{z}_n = \nu_n \, \mathbb{Z} \, \mathbf{z}_n, \qquad n = 1, \ldots, N. \tag{14}$$

Proof. Since $\sup_{\eta \in \mathcal{U}_M} \frac{(\eta, z)}{\|\eta\|} = \|\Pi_{\mathcal{U}_M} z\|$, we obtain:

$$\beta_{N,M}^2 \; = \; \inf_{z \in \mathcal{Z}_N} \sup_{\eta \in \mathcal{U}_M} \left(\frac{(\eta, z)}{\|\eta\| \|z\|} \right)^2 \; = \; \inf_{z \in \mathcal{Z}_N} \; \frac{\|\Pi_{\mathcal{U}_M} z\|^2}{\|z\|^2}.$$

We observe that for any $z \in \mathcal{Z}_N$ the projection onto $\mathcal{U}_M$ can be written as $\Pi_{\mathcal{U}_M} z = \sum_{m=1}^M \eta_m^z \, K_{x_m}$, where the vector $\boldsymbol{\eta}^z$ satisfies $\boldsymbol{\eta}^z = \mathbb{K}^{-1} \, \mathbb{L} \, \mathbf{z}$. Therefore, we find

$$\beta_{N,M}^2 \; = \; \inf_{\mathbf{z} \in \mathbb{R}^N} \; \frac{\mathbf{z}^T \, \mathbb{L}^T \, \mathbb{K}^{-1} \, \mathbb{L} \, \mathbf{z}}{\mathbf{z}^T \mathbb{Z} \mathbf{z}}.$$

Introducing the Lagrangian multiplier $\nu \in \mathbb{R}$, we can write the optimality conditions as

$$\begin{cases} \mathbb{L}^T \mathbb{K}^{-1} \mathbb{L}\, \mathbf{z} - \nu \mathbb{Z}\mathbf{z} & = \mathbf{0} \\ \mathbf{z}^T \mathbb{Z}\mathbf{z} & = 1 \end{cases}$$

Thesis follows. □

Proof. (Proposition 2.5) It is straightforward to observe that we can restate the minimization statement (2) as

$$J_\xi(u_\xi^\star) = \xi\, \boldsymbol{\eta}^T \mathbb{K}\, \boldsymbol{\eta} + \frac{1}{M} \|\mathbb{K}\boldsymbol{\eta} + \mathbb{L}\boldsymbol{z} - \mathbf{y}\|_2^2, \qquad \text{subject to } \mathbb{L}^T \boldsymbol{\eta} = \mathbf{0},$$

where the constraint imposes that $\eta_\xi^\star \in \mathcal{Z}_N^\perp$. Then, by introducing the Lagrangian multiplier $\boldsymbol{\upsilon} \in \mathbb{R}^N$, we can derive the optimality system:

$$\begin{cases} \left(\xi\mathbb{K} + \frac{1}{M}\mathbb{K}^2\right) \boldsymbol{\eta}^\star + \frac{1}{M}\mathbb{K}\mathbb{L}\mathbf{z}^\star + \mathbb{L}\,\boldsymbol{\upsilon} & = \frac{1}{M}\mathbb{K}\mathbf{y} \\ \mathbb{L}^T \mathbb{K}\, \boldsymbol{\eta}^\star + \mathbb{L}^T\mathbb{L}\mathbf{z}^\star & = \mathbb{L}^T\mathbf{y} \\ \mathbb{L}^T \boldsymbol{\eta}^\star & = \mathbf{0} \end{cases} \tag{15}$$

By pre-multiplying $(15)_1$ by $\mathbb{L}^T\mathbb{K}^{-1}$, we obtain

$$\mathbb{L}^T \mathbb{K}\, \boldsymbol{\eta}^\star + \mathbb{L}^T\mathbb{L}\mathbf{z}^\star + M\mathbb{L}^T\mathbb{K}^{-1}\mathbb{L}\,\boldsymbol{\upsilon} = \mathbb{L}^T\mathbf{y}.$$

Since $\mathbb{L}^T\mathbb{K}^{-1}\mathbb{L}$ is full-rank (it follows from the assumption $\beta_{N,M} > 0$ and Lemma 2.6), by comparing with $(15)_2$, we must have $\boldsymbol{\upsilon} = \mathbf{0}$. As a result, we can restate (15), as

$$\begin{cases} (\xi M\mathbb{I} + \mathbb{K})\, \boldsymbol{\eta}^\star + \mathbb{L}\mathbf{z}^\star & = \mathbf{y} \\ \mathbb{L}^T \mathbb{K}\, \boldsymbol{\eta}^\star + \mathbb{L}^T\mathbb{L}\mathbf{z}^\star & = \mathbb{L}^T\mathbf{y} \\ \mathbb{L}^T \boldsymbol{\eta}^\star & = \mathbf{0} \end{cases} \tag{16}$$

Finally, we observe that $(16)_2$ follows from $(16)_1$ and $(16)_3$. We have indeed

$$\mathbb{L}^T (\mathbb{K}\boldsymbol{\eta}^\star + \mathbb{L}\boldsymbol{z}^\star - \mathbf{y}) = \mathbb{L}^T ((\mathbb{K} + \xi M\mathbb{I})\, \boldsymbol{\eta}^\star + \mathbb{L}\boldsymbol{z}^\star - \mathbf{y}) = \mathbf{0}.$$

Thesis follows. □

We now state an important corollary. Proof is straightforward and is here omitted.

Corollary 2.7. *Problem* (2) *admits an unique solution of the form* (12) *if and only if the matrix* $\mathbb{L}$ *has rank equal to* N.

Some comments are in order. First, Corollary 2.7 corresponds to the original result proved by Kimeldorf and Wahba ([16, Theorem 5.1]). Second, Corollary 2.7 provides an actionable condition to verify the well-posedness of Problem (2). We further observe that this condition does not rely on the assumption that the background belongs to $\mathcal{U}$. This observation motivates the next result.

Proposition 2.8. *Suppose that* $\mathcal{Z}_N \subset C(\Omega)$ *and let rank*$(\mathbb{L}) = N$. *Then, the solution to* (10) *exists, is unique and is of the form* (12) *with* $(\boldsymbol{\eta}^\star, \mathbf{z}^\star) \in \mathbb{R}^M \times \mathbb{R}^N$ *satisfying* (13).

Proof. We denote by $\hat{\eta}_\xi$, $\hat{z}_\xi$ the solution to (10). We first observe that

$$V_M(\hat{\eta}_\xi + \hat{z}_\xi) = V_M(\Pi_{\mathcal{U}_M}\hat{\eta}_\xi + \hat{z}_\xi).$$

Therefore, any solution to (10) is of the form (12). Substituting in the minimization statement, we obtain the following algebraic minimization problem for $(\boldsymbol{\eta}^\star, \mathbf{z}^\star) \in \mathbb{R}^M \times \mathbb{R}^N$

$$\min_{(\boldsymbol{\eta}^\star, \boldsymbol{z}^\star)\in\mathbb{R}^M\times\mathbb{R}^N} \xi\, \boldsymbol{\eta}^T\, \mathbb{K}\, \boldsymbol{\eta} + \frac{1}{M}\, \|\mathbb{K}\boldsymbol{\eta} + \mathbb{L}\boldsymbol{z} - \mathbf{y}\|_2^2.$$

By deriving the stationary conditions, we find

$$\begin{cases} \left(\xi\mathbb{K} + \frac{1}{M}\mathbb{K}^2\right)\, \boldsymbol{\eta}^\star \,+\, \frac{1}{M}\mathbb{K}\mathbb{L}\mathbf{z}^\star & = \frac{1}{M}\mathbb{K}\mathbf{y} \\ \mathbb{L}^T\, \mathbb{K}\, \boldsymbol{\eta}^\star + \mathbb{L}^T\mathbb{L}\mathbf{z}^\star & = \mathbb{L}^T\mathbf{y} \end{cases} \tag{17}$$

By premultiplying $(17)_1$ by $M\mathbb{K}^{-1}$, we find

$$(\xi M\mathbb{I} + \mathbb{K})\, \boldsymbol{\eta}^\star + \mathbb{L}\mathbf{z}^\star = \mathbf{y}. \tag{18a}$$

If we now premultiply the latter equation by $\mathbb{L}^T$ and we subtract $(17)_2$, we obtain

$$\mathbb{L}^T\boldsymbol{\eta} = \mathbf{0}. \tag{18b}$$

Saddle system (18a) - (18b) is well-posed since $\mathbb{K}$ is invertible and $\mathbb{L}$ is full-rank by hypothesis. Thesis follows. □

2.5. **Construction of the spaces**

2.5.1. *Background spaces $\mathcal{Z}_N$*

In many engineering applications, mathematical models -which in mechanics typically consist of partial differential equations (PDEs) - have been developed to describe the behavior of the physical state; these models typically depend on a set of parameters $\mu \in \mathbb{R}^P$ representing material properties, geometry, operating conditions etc. In practical applications, such parameters are uncertain and thus only a confidence region $\mathcal{P} \subset \mathbb{R}^P$ is available. As a result, provided that the model is reasonably accurate, the true field u^{true} is close (in a suitable functional norm) to the solution $u^{bk}(\mu^\star)$ to the parametrized best-knowledge model for some unknown parameter $\mu^\star \in \mathcal{P}$.

Following the idea proposed in [24], in this work, we define the background space $\mathcal{Z}_N$ such that for each parameter $\mu \in \mathcal{P}$ the solution to the PDE model $u^{bk}(\mu)$ can be accurately represented by an element of the space. As a result, assuming that $u^{true} \approx u^{bk}(\mu^\star)$ for some $\mu^\star \in \mathcal{P}$, then there exists an unknown element $\hat{z}_N \in \mathcal{Z}_N$ such that $u^{true} \approx \hat{z}_N$.

The problem of determining a low-dimensional space $\mathcal{Z}_N$ that accurately represents the PDE solution manifold $\{u^{bk}(\mu) : \mu \in \mathcal{P}\}$ has been extensively studied in the last few decades in the Model Order Reduction literature. The choice of the particular algorithm to build $\mathcal{Z}_N$ depends on the structure of the equation and on the parameterization. In this work, following [24], we employ the weak-Greedy algorithm ([28, Section 7.2.2]). The method properly selects N parameters $\{\mu_n^\star\}_{n=1}^N \subset \mathcal{P}$ and sets $\mathcal{Z}_N = \text{span}\{u^{bk}(\mu_n^\star)\}_{n=1}^N$.

Some comments are in order. In our framework, the mathematical model is only used to build the background space $\mathcal{Z}_N$ and is not directly employed during the data assimilation procedure. In addition, the PBDW statement does not depend on how the space $\mathcal{Z}_N$ is built. These two observations provide us with some flexibility in choosing the background space. We could for instance first define the space $\mathcal{Z}_N$ using a parameterized best-knowledge model, and then augment the space exploiting historical data. The latter observation could be attractive for applications in which reliable historical data are available. In this work, we do not pursue this feature of our approach; we refer to [25] for a detailed discussion about this issue.

2.5.2. *Choice of the Reproducing Kernel*

In approximation theory, several choices of the kernel function K have been proposed and analyzed both from the theoretical and the numerical perspective. We refer to [34] and to the references therein for a detailed discussion. Here, we only introduce the class of kernels employed in the numerical simulations.

In this work, we consider compactly supported radial basis functions of minimal degree (csRBFs, [34]), also known as Wendland functions. This class of kernels is defined as $K_\chi(x,y) = \phi_{d,k}(\gamma\|x-y\|_2)$ where $\chi = [k,\gamma]$ and

$$\phi_{d,k}(r) = \begin{cases} p_{d,k}(r) & 0 \le r \le 1; \\ 0 & r > 1. \end{cases} \tag{19a}$$

The polynomial $p_{d,k}$ has the following form for $k = 0, 1$ and for all d:

$$p_{d,k}(r) = \begin{cases} (1-r)^{\ell_{d,k}} & k = 0 \\ (1-r)^{\ell_{d,k}+1}\left((\ell_{d,k}+1)r+1\right) & k = 1 \end{cases} \tag{19b}$$

and $\ell_{d,k} = \left\lfloor \frac{d}{2} \right\rfloor + k + 1$. We observe that it is possible to generalize (19b) to the more general case $k \in \mathbb{N}$; we refer to [34, Table 9.1] for the explicit formulas.

The next result clarifies the connection between csRBF and Sobolev spaces. We refer to [34, Theorem 10.35] for the proof.

Proposition 2.9. *Let us consider the compactly supported RBF K_χ, $K_\chi(x,y) = \phi_{d,k}(\gamma\|x-y\|_2)$, introduced in* (19)*. Let $\Omega = \mathbb{R}^d$, and let either one of these conditions hold:*

(1) $d \ge 3$, $k \ge 0$;
(2) $d \ge 1$, $k > 0$.

Then, the native space for K_χ is the Sobolev space $H^{(d+1)/2+k}(\mathbb{R}^d)$.

Some comments are in order. By restricting ourselves to csRBF kernels, the choice of the inner product reduces to the choice of the parameters $\chi = [k,\gamma]$. We refer to section 4 for a thorough discussion about their choice. We further observe that, by resorting to these kernels, we lose the possibility of imposing strong (Dirichlet) boundary conditions at the boundary of Ω in the variational formulation.

2.6. Connection with other formulations

2.6.1. *Connections with deterministic formulations*

For suitable choices of the hyper-parameters, our formulation reduces to the Generalized Empirical Interpolation Method (GEIM, [23]), to the original PBDW formulation presented in [24] and to least-square regression (LSR). If we choose $N = M$, then $u^\star = z^\star$ corresponds to the solution to the GEIM formulation for pointwise measurements. To prove the other two equivalences, we first introduce the formulations:

(1) (PBDW) find $u^\star = \eta^\star + z^\star$ such that

$$(\eta^\star,\, z^\star) := \arg \min_{(\eta,z)\in\mathcal{U}\times\mathcal{Z}_N} \|\eta\| \qquad \text{subject to } \eta(x_m) + z(x_m) = y_m,\ \ m = 1,\dots,M; \tag{20}$$

(2) (LSR) find $z^\star_{LS} \in \mathcal{Z}_N$ such that

$$z^\star_{LS} := \arg\min_{z\in\mathcal{Z}_N} V_M(z). \tag{21}$$

When $\mathcal{Z}_N$ is built using a Proper Orthogonal Decomposition (POD, [19]), statement (21) corresponds to Gappy-POD ([11, 35]).

We now show the equivalence.

Proposition 2.10. *Let $\beta_{N,M} > 0$. Let $u^\star_\xi = \eta^\star_\xi + z^\star_\xi$ be the solution to (2). Then, we have*

$$\lim_{\xi \to 0^+} \|u^\star - u^\star_\xi\| = 0, \tag{22}$$

and

$$\lim_{\xi \to \infty} \|u^\star_\xi - z^\star_{LS}\| = 0. \tag{23}$$

Proof. We first derive algebraic formulations for the solutions to (20) and (21). Recalling [24, Section 2.4], we have that $u^\star$ is of the form (12) with coefficients $\boldsymbol{\eta}^\star_0, \boldsymbol{z}^\star_0$ satisfying

$$\begin{bmatrix} \mathbb{K} & \mathbb{L} \\ \mathbb{L}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}^\star_0 \\ \mathbf{z}^\star_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \mathbf{0} \end{bmatrix},$$

and thus

$$\boldsymbol{z}^\star_0 = \left(\mathbb{L}^T \mathbb{K}^{-1} \mathbb{L}\right)^{-1} \mathbb{L}^T \mathbb{K}^{-1} \boldsymbol{y}; \qquad \boldsymbol{\eta}^\star_0 = \mathbb{K}^{-1} \left(\mathbb{L} \mathbf{z}^\star_0 - \boldsymbol{y}\right).$$

Similarly, the solution to (21) can also be written in the form (12) with

$$\boldsymbol{z}^\star_\infty = \left(\mathbb{L}^T \mathbb{L}\right)^{-1} \mathbb{L}^T \boldsymbol{y}; \qquad \boldsymbol{\eta}^\star_\infty = \mathbf{0}.$$

Exploiting Proposition 2.5, if we denote for convenience by $\boldsymbol{\eta}^\star_\xi$ and $\boldsymbol{z}^\star_\xi$ the vectors of coefficients associated with $u^\star_\xi$, we need to show that

$$\lim_{\xi \to 0^+} (\boldsymbol{\eta}^\star_\xi, \boldsymbol{z}^\star_\xi) = (\boldsymbol{\eta}^\star_0, \boldsymbol{z}^\star_0) \qquad \lim_{\xi \to \infty} (\boldsymbol{\eta}^\star_\xi, \boldsymbol{z}^\star_\xi) = (\boldsymbol{\eta}^\star_\infty, \boldsymbol{z}^\star_\infty).$$

The proofs of these two limits are straightforward and are here omitted. □

2.6.2. *Connection with 3D-VAR*

To provide insights about the formulation and its intimate connection with 3D-VAR, we derive APBDW starting from 3D-VAR exploiting a probabilistic interpretation. With this in mind, we define the 3D-VAR statement in a variational fashion ([4, Chapter 2]): find $u^\star \in \mathcal{U}$ such that

$$u^\star := \arg\min_{u \in \mathcal{U}} \frac{1}{2} \|u - u^{bk}\|^2 + \frac{1}{2} \|\mathcal{L}_M(u) - \boldsymbol{y}\|^2_{\mathbb{W}};$$

where u^{bk} is the non-parametric background, $\mathcal{L}_M(u) = (u(x_1), \ldots, u(x_M))$, $\mathbb{W}$ is the observation error covariance, and $\|\boldsymbol{d}\|_{\mathbb{W}} = \sqrt{\boldsymbol{d}^T \mathbb{W} \boldsymbol{d}}$. We observe that if we substitute u^{bk} with the parameterized background $u^{bk}(\mu)$, $\mu \in \mathcal{P}$, we obtain the so-called partial spline model ([33, Chapter 9]): find $(\mu^\star, \eta^\star) \in \mathcal{P} \times \mathcal{U}$ such that

$$(\mu^\star, u^\star) = \arg \min_{(\mu,u) \in \mathcal{P} \times \mathcal{U}} \frac{1}{2} \|u - u^{bk}(\mu)\|^2 + \frac{1}{2} \|\mathcal{L}_M(u) - \boldsymbol{y}\|^2_{\mathbb{W}};$$

or, equivalently,

$$(\mu^\star, \eta^\star) = \arg \min_{(\mu,\eta) \in \mathcal{P} \times \mathcal{U}} \frac{1}{2} \|\eta\|^2 + \frac{1}{2} \|\mathcal{L}_M(u^{bk}(\mu) + \eta) - \boldsymbol{y}\|^2_{\mathbb{W}}, \tag{24}$$

with $u^\star = u^{bk}(\mu^\star) + \eta^\star$. If we introduce the rank-N approximation ([7]) of the best-knowledge field $u^{bk}(\mu)$,

$$u^{bk}_N(x; \mu) = \sum_{n=1}^{N} \phi_n(\mu)\, \zeta_n(x), \qquad x \in \Omega, \;\; \mu \in \mathcal{P},$$

we can approximate statement (24) as follows:

$$(\mu^\star, \eta^\star) = \arg \min_{(\mu,\eta)\in\mathcal{P}\times\mathcal{U}} \frac{1}{2}\|\eta\|^2 + \frac{1}{2}\big\|\mathcal{L}_M\big(\sum_{n=1}^{N} \phi_n(\mu)\,\zeta_n + \eta\big) - \boldsymbol{y}\big\|_{\mathbb{W}}^2. \tag{25}$$

We can now relax problem (25) as follows:

$$(\boldsymbol{\phi}^\star, \eta^\star) = \arg \min_{(\boldsymbol{\phi},\eta)\in\mathbb{R}^N\times\mathcal{U}} \frac{1}{2}\|\eta\|^2 + \frac{1}{2}\big\|\mathcal{L}_M\big(\sum_{n=1}^{N} \phi_n\,\zeta_n + \eta\big) - \boldsymbol{y}\big\|_{\mathbb{W}}^2,$$

which can also be written as

$$(z^\star, \eta^\star) = \arg \min_{(z,\eta)\in\mathcal{Z}_N\times\mathcal{U}} \frac{1}{2}\|\eta\|^2 + \big\|\mathcal{L}_M\big(z + \eta\big) - \boldsymbol{y}\big\|_{\mathbb{W}}^2,$$

where $\mathcal{Z}_N := \text{span}\{\zeta_n\}_{n=1}^N$. The latter corresponds to the two-field APBDW formulation (10) for $\mathbb{W} = \frac{1}{\xi M}\mathbb{I}$.

We provide some remarks. First, our derivation allows us to interpret APBDW as a convex relaxation of the partial spline model for a *parametric affine* background. Instead of penalizing the difference between the state estimate and the manifold $\mathcal{M}^{bk} = \{u^{bk}(\mu) : \mu \in \mathcal{P}\}$, we penalize the distance from the linear space $\mathcal{Z}_N$. This derivation motivates the interpretation of $z_\xi^\star$ as the deduced background estimate, the component of the state informed by the prior knowledge of the system, and the interpretation of $\eta_\xi^\star$ as update. Second, this derivation allows to interpret APBDW as a Gaussian linear system with an improper prior. This interpretation has been first proposed by Wahba in [32]. In this work, we do not exploit this interpretation to develop credible intervals for the state estimation error.

We now comment on the relaxation process. If we denote by $\bar{\mu} \in \mathcal{P}$ the centroid of the parameter space and we normalize the directions $\zeta_1, \ldots, \zeta_N$, we typically observe that

$$\max_{\mu\in\mathcal{P}} |\phi_1(\mu) - \phi_1(\bar{\mu})| \gg \max_{\mu\in\mathcal{P}} |\phi_2(\mu) - \phi_2(\bar{\mu})| \gg \ldots$$

The relaxation step that leads to APBDW discards this information associated with the best-knowledge map. Despite the gain in computational efficiency, this relaxation might lead to a deterioration in performance. Extension of APBDW in this direction is subject of ongoing research. In this respect, we mention the work of Binev et al. ([5]) that proposed the multi-space problem to take into account the anisotropy in the coefficients of the expansion.

3. Error analysis

We present *a priori* and *a posteriori* estimates for the $L^2(\Omega)$ state-estimation error $\|u^{true} - u_\xi^\star\|_{L^2(\Omega)}$. The importance of the error analysis is twofold. First, it motivates our formulation from a theoretical viewpoint. Second, it provides insights about the role of the different pieces of our formulation: the regularization parameter ξ, the background space $\mathcal{Z}_N$, the kernel K and the centers $\mathcal{X}_M$.

3.1. *A priori* error analysis

In order to derive error bounds for the $L^2(\Omega)$ state-estimation error $\|u^{true} - u_\xi^\star\|_{L^2(\Omega)}$, we must first introduce assumptions on our dataset $\mathcal{D}_M$. To our knowledge, three different scenarios have been considered so far.

(1) *Random-design regression*: the pairs $\{(x_m, y_m)\}_{m=1}^M$ are drawn independently from a joint unknown distribution $\rho_{(X,Y)}$. In this case, the objective of learning is to estimate the conditional expectation $\mathbb{E}[Y|X = x]$.

(2) *Fixed-design regression*: the centers $\mathcal{X}_M = \{x_1, \dots, x_M\}$ are fixed (non-random) points in Ω, while the responses $\mathcal{Y}_M = \{y_m\}_{m=1}^M$ satisfy $y_m = u^{true}(x_m) + \epsilon_m$, where $u^{true} : \Omega \mapsto \mathbb{R}$ is the deterministic field of interest and $\epsilon_1, \dots, \epsilon_M$ are independent identically distributed (i.i.d.) random variables with zero mean and variance σ^2, $\epsilon_m \sim (0, \sigma^2)$.

(3) *Scattered data approximation*: both centers $\mathcal{X}_M$ and responses $\mathcal{Y}_M$ are non-random, and we assume that there exists some unknown $\delta > 0$ such that $|y_m - u^{true}(x_m)| \leq \delta$ for all $m = 1, \dots, M$.

The first scenario has been extensively studied in the statistical learning literature (see, e.g., [9, 30]). We refer to [13] for a complete review of the error bounds available. The second scenario has also been studied in statistics; we refer to the survey [12] for further details about a specific class of kernels. Finally, the third scenario has been studied in approximation theory and radial basis functions (see, e.g., [34]). From the modeling perspective, the first scenario refers to the case in which we do not have control on the observation centers, the second scenario addresses the problem of random error in the measurements, and the third scenario addresses the problem of systematic deterministic error.

In the next two sections, we present error bounds for both the second and the third scenarios. Our analysis for the third scenario is largely inspired by the work of Krebs, Louis and Wendland ([18]). On the other hand, the analysis for fixed-design regression seems new. We state upfront that in the remainder of this section we assume that $\mathcal{Z}_N \subset \mathcal{U}$.

3.1.1. *An* a priori *error bound for scattered data approximation*

We state the main result of this section.

Proposition 3.1. *Let Ω be a Lipschitz domain and let $\mathcal{U}$ be the Sobolev space $H^\tau(\Omega)$ with $\tau > d/2$. Let $\beta_{N,M}$ in (3) be strictly positive.*

Then, if $u^{true} \in \mathcal{U}$, the following holds:

$$\|u^{true} - u_\xi^\star\|_{L^2(\Omega)}^2 \leq C_{N,\mathcal{X}_M} \left(h_{\mathcal{X}_M}^{2\tau} \left(2\|\Pi_{\mathcal{Z}_N^\perp} u^{true}\| + \frac{\delta}{2} \frac{1}{\sqrt{\xi}} \right)^2 + h_{\mathcal{X}_M}^d M \left(\delta + \frac{\sqrt{\xi}}{2} \|\Pi_{\mathcal{Z}_N^\perp} u^{true}\| \right)^2 \right), \tag{26a}$$

where $C_{N,\mathcal{X}_M}$ is defined as

$$C_{N,\mathcal{X}_M} := \sup_{u \in \mathcal{U}} \frac{\|u\|_{L^2(\Omega)}^2}{h_{\mathcal{X}_M}^{2\tau} \|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + h_{\mathcal{X}_M}^d \|u\|_{\ell^2(\mathcal{X}_M)}^2}, \tag{26b}$$

$\|u\|_{\ell^2(\mathcal{X}_M)} := \sqrt{\sum_{m=1}^M u(x_m)^2}$*, and the fill distance $h_{\mathcal{X}_M}$ as*

$$h_{\mathcal{X}_M} = \sup_{x \in \Omega} \min_{m=1,\dots,M} \|x - x_m\|_2. \tag{26c}$$

Proof of Proposition 3.1 is technical and for this reason we present it in Appendix A. In the remainder of this section, we state a number of remarks.

Remark 3.2. We observe that the constant $C_{N,\mathcal{X}_M}$ is associated with the maximum eigenvalue associated to a generalized eigenproblem. Provided that the inf-sup constant $\beta_{N,M} > 0$ and $h_{\mathcal{X}_M} < 1$, the constant $C_{N,\mathcal{X}_M}$ defined in (26b) can be estimated as follows:

$$C_{N,\mathcal{X}_M} \leq \frac{1}{\min\{c_{N,M}, 1 - h_{\mathcal{X}_M}^{2\tau - d}\}} C, \tag{27}$$

where $c_{N,M}$ is defined in (6). We rigorously prove (27) in Appendix A. We remark that to practically estimate $C_{N,\mathcal{X}_M}$, we need to numerically approximate the maximum eigenvalue of the generalized eigenproblem associated with $C_{N,\mathcal{X}_M}$. We refer to [14] for a discussion on the use of meshless methods based on csRBF for the solution to eigenproblems.

Remark 3.3. For quasi-uniform grids, $h_{\mathcal{X}_M} \sim M^{-1/d}$, for $M \to \infty$, the right-hand side reduces to

$$\|u^{true} - u_\xi^\star\|^2_{L^2(\Omega)} \lesssim \mathcal{O}\left(\|\Pi_{\mathcal{Z}_N^\perp} u^{true}\|^2 \, h^{2\tau}_{\mathcal{X}_M} \left(1 + \frac{1}{\Lambda}\right)^2 + \delta^2 \, (1+\Lambda)^2\right) \tag{28a}$$

where

$$\Lambda = \frac{\sqrt{\xi}\|\Pi_{\mathcal{Z}_N^\perp} u^{true}\|}{\delta}, \tag{28b}$$

By minimizing with respect to Λ, we obtain that the asymptotically optimal choice of ξ is

$$\xi = \left(\frac{\delta}{\|\Pi_{\mathcal{Z}_N^\perp} u^{true}\|}\right)^{2/3} h^{4/3\tau}_{\mathcal{X}_M}. \tag{29a}$$

For this choice of the hyper-parameter, we obtain:

$$\|u^{true} - u_\xi^\star\|^2_{L^2(\Omega)} \lesssim \mathcal{O}\left(\|\Pi_{\mathcal{Z}_N^\perp} u^{true}\|^{2/3} \, h^{2/3\tau}_{\mathcal{X}_M} \, \delta^{4/3} + \delta^2\right) \qquad M \to \infty. \tag{29b}$$

We observe that for any finite $\delta > 0$, we do not expect convergence in a L^2 sense. We also observe that the optimal value of ξ is directly proportional to δ, inversely proportional to the background best-fit error $\|\Pi_{\mathcal{Z}_N^\perp} u^{true}\|$ and decreases as M increases.

Remark 3.4. In the case of perfect measurements, estimate (26) reduces to

$$\|u^{true} - u_\xi^\star\|^2_{L^2(\Omega)} \leq \frac{1}{4} C_{N,\mathcal{X}_M} \left(16 h^{2\tau}_{\mathcal{X}_M} + h^d_{\mathcal{X}_M} M\xi\right) \|\Pi_{\mathcal{Z}_N^\perp} u^{true}\|^2. \tag{30}$$

If we neglect the factor $C_{N,\mathcal{X}_M}$, we observe a multiplicative effect between M convergence (associated with the update) and N convergence (associated with the deduced background).

3.1.2. A priori *error bounds for fixed-design regression*

We first introduce some notation. First, we define the matrix $\mathbb{A}_\xi \in \mathbb{R}^{N+M,N+M}$,

$$\mathbb{A}_\xi := \begin{bmatrix} \xi M \mathbb{I}_M + \mathbb{K} & \mathbb{L} \\ \mathbb{L}^T & 0 \end{bmatrix}, \tag{31a}$$

associated with the linear system (13). Then, we introduce $\Sigma \in \mathbb{R}^{M,M}$,

$$\Sigma := \begin{bmatrix} \mathbb{I}_M & 0 \\ 0 & 0 \end{bmatrix}. \tag{31b}$$

Finally, we introduce $\mathbb{M} \in \mathbb{R}^{N+M,N+M}$ such that

$$\mathbb{M}_{i,i'} := \int_\Omega \psi_i(x)\, \psi_{i'}(x)\, dx, \qquad \psi_i(x) = \begin{cases} K_{x_i} & i = 1, \ldots, M \\ \zeta_{i-M} & i = M+1, \ldots, M+N \end{cases} \tag{31c}$$

We can now state the error bound.

Proposition 3.5. *Let Ω be a Lipschitz domain and let $\mathcal{U}$ be the Sobolev space $H^\tau(\Omega)$ with $\tau > d/2$. Let $\beta_{N,M}$ in* (3) *be strictly positive.*

Then, if $u^{true} \in \mathcal{U}$, the following holds:

$$\mathbb{E}\left[\|u^{true} - u_\xi^\star\|^2_{L^2(\Omega)}\right] \leq \frac{1}{2} C_{N,\mathcal{X}_M} \left(16 h^{2\tau}_{\mathcal{X}_M} + h^d_{\mathcal{X}_M} M\xi\right) \|\Pi_{\mathcal{Z}_N^\perp} u^{true}\|^2 + 2\sigma^2 \, trace\left(\mathbb{A}_\xi^{-1} \mathbb{M} \mathbb{A}_\xi^{-1} \Sigma\right). \tag{32}$$

Proof of Proposition 3.5 is presented in Appendix A. We observe that, unlike in the previous case, it is not evident how to provide explicit estimates for the optimal value of ξ. We also observe that (32) can be easily extended to correlated noise.

3.2. *A posteriori* error analysis

Next result provides the identity of interest.

Proposition 3.6. *Let $\{x_i\}_{i=1}^I$ be drawn independently from an uniform distribution over Ω. Let $y_i = u^{true}(x_i) + \delta_i + \epsilon_i$, where $\epsilon_1, \dots \epsilon_I$ are i.i.d. random variables such that $\epsilon_i \sim (0, \sigma^2)$ and $\delta_1, \dots, \delta_I$ are deterministic unknown disturbances. Let us further assume that $\{x_i\}_{i=1}^I$ and $\{\epsilon_i\}_{i=1}^I$ are independent random sequences.*

Then, we have that the mean squared error

$$MSE_I := \frac{1}{I} \sum_{i=1}^I \left(y_i - u_\xi^\star(x_i)\right)^2 \tag{33}$$

satisfies

$$\mathbb{E}[MSE_I] = E^2_{mean} + \sigma^2 + \frac{1}{I} \sum_{i=1}^I \delta_i^2 - \frac{2}{|\Omega| I} \sum_{i=1}^I \delta_i \left(\int_\Omega u^{true}(x) - u_\xi^\star(x)\, dx\right) \tag{34}$$

where E^2_{mean} is defined as follows:

$$E^2_{mean} := \frac{1}{|\Omega|} \int_\Omega \left(u(\mu) - u_\xi^\star\right)^2 dx. \tag{35}$$

Proof. To simplify notation, we introduce the random sequence $\{e_i = u^{true}(x_i) - u_\xi^\star(x_i)\}_{i=1}^I$. We observe that $e_1, \dots, e_I$ are i.i.d. and $\mathbb{E}[e_i^2] = \frac{1}{|\Omega|} \|u^{true} - u^\star_{\lambda,\xi}\|^2_{L^2(\Omega)}$. Then, exploiting linearity of the expected value operator and the fact that $\{x_i\}_{i=1}^I$ and $\{\epsilon_i\}_{i=1}^I$ are independent, we find

$$\mathbb{E}[MSE_I] = \mathbb{E}\left[e_1^2\right] + \mathbb{E}\left[\epsilon_1^2\right] + \frac{1}{I} \sum_{i=1}^I \delta_i^2 - \frac{2}{I} \sum_{i=1}^I \delta_i \, \mathbb{E}[e_i].$$

Thesis follows. □

In absence of systematic noise ($\delta_i \equiv 0$), identity (34) reduces to

$$\mathbb{E}[MSE_I] = E^2_{mean} + \sigma^2. \tag{36}$$

Estimate (36) shows that for random noise ($\delta_i \equiv 0$) the mean squared error (33) can be used to asymptotically bound the squared $L^2(\Omega)$ error. Furthermore, since σ^2 is independent of the state estimate, minimizing the mean squared error is equivalent to minimize the $L^2(\Omega)$ error. The latter observation motivates the adaptive strategy presented in section 4.

4. Model adaptation

As observed in the previous sections, our procedure depends on a fair amount of design choices, which include the choice of a number of hyper-parameters and the choice of the observation centers and background space $\mathcal{Z}_N$. In section 4.1, we discuss how to exploit the error analysis to perform some design choices *a priori*. Then, in section 4.2, we discuss the adaptive strategy used to tune the parameters of the formulation after having acquired data.

4.1. *A priori* considerations

We first recall the APBDW formulation and we discuss separately each element of the formulation. The APBDW state estimate $u_\xi^\star$ is given by

$$u_\xi^\star := \arg\min_{u\in\mathcal{U}} \xi\|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + \frac{1}{M}\sum_{m=1}^{M} \left(u(x_m) - y_m\right)^2 \tag{37}$$

where $\mathcal{Z}_N = \text{span}\{\zeta_n\}_{n=1}^N$. We observe that the formulation depends on the regularization parameter ξ, the sensor locations $\mathcal{X}_M = \{x_m\}_{m=1}^M$ and the choice of the RKHS norm $\|\cdot\|$. Recalling Moore-Aronszajn theorem, the latter choice is equivalent to the choice of the reproducing kernel K.

The hyper-parameter $\xi > 0$ controls the amount of regularization introduced: for $\xi = 0$, the solution to (37) interpolates exactly the data while for $\xi \to \infty$, the solution to (37) converges to the least-squares solution. Our error analysis shows that the choice of ξ strongly depends on the noise variance and on the maximum systematic error δ; both these quantities are hard to estimate *a priori*.

We now discuss the choice of the kernel K. Since in this work we employ csRBF kernels, this reduces to the choice of the hyper-parameters k and γ in $K_\chi(x,y) = \phi_{d,k}(\gamma\|x-y\|_2)$, where $\phi_{d,k}$ is defined in (19). As stated in Proposition 2.9, the parameter k determines the Sobolev regularity of the RKHS. Since in practical applications we expect that measurements are noisy, recalling estimate (29), we choose the minimal value of k for which K is a reproducing kernel, which is $k = 1$ for $d = 1, 2$ and $k = 0$ for $d \geq 3$. The parameter γ regulates the length scale of the kernel functions. In our experience, for small values of M, the choice of γ weakly influences the results; we can thus pick γ *a priori* such that the kernel functions K_{x_m} share the same length scale with the elements of $\mathcal{Z}_N$. On the other hand, for larger values of M, the choice of γ significantly influences the performances of the method and it must be adapted using data. We remark that by changing γ we effectively modify the inner product $(\cdot,\cdot)$.

We comment on the choice of the sensor locations. If we neglect the effect of the sensor locations on the stability constant $C_{N,\mathcal{X}_M}$, the error analysis suggests to choose the observation centers to minimize the fill distance $h_{\mathcal{X}_M}$ in (26c). For $N \simeq M$, sensor location might influence significantly the value of $C_{N,\mathcal{X}_M}$. As a result, it might be worth to choose the observation centers to maximize $C_{N,\mathcal{X}_M}$ for any given M. For the PBDW formulation, in [24], the authors propose a Greedy strategy for the selection of the observation centers to maximize the inf-sup constant $\beta_{N,M}$ defined in (3). In this work, we simply consider equispaced observation centers and we refer to a future work for more elaborated strategies for the selection of the observation centers that address both stability and approximation.

Motivated by the previous considerations, in our numerical simulations, we choose adaptively the regularization parameter ξ and the kernel parameter γ. In the next section, we present the algorithm used to perform online adaptation.

4.2. Adaptive procedure

In the Statistical Learning literature, several approaches have been presented to tune the design parameters of reguralized regression formulations; we refer to [15, Chapter 7] and to [17] for a thorough overview. The adaptive strategy depends on the size of the dataset, which in our context corresponds to the amount of available transducers. If we denote by L the number of available transducers and by $\mathcal{D}_L = \{(x_\ell, y_\ell)\}_{\ell=1}^L$ the corresponding

dataset, for large values of L, the holdout method is the most widely used approach. On the other hand, for small values of L, κ-fold cross-validation is typically employed. In the remainder of this section, we briefly review these techniques and we discuss their application to our problem.

The holdout method partitions the dataset $\mathcal{D}_L$ into the two mutually exclusive subsets $\mathcal{D}_M = \{(x_m, y_m)\}_{m=1}^M$ and $\mathcal{D}_I = \{(x_i, y_i)\}_{i=1}^I$. Given the finite dimensional search space $\mathfrak{H}^{hyper}$ for (ξ, γ), we generate the state estimate $u^\star_{\xi,\gamma}$ based on the training set and then we compute the mean squared error over the validation set

$$MSE_I(\xi, \gamma) = \frac{1}{I} \sum_{i=1}^{I} \left(y_i - u^\star_{\xi,\gamma}(x_i) \right)^2, \tag{38}$$

for each (ξ, γ) in $\mathfrak{H}^{hyper}$. Finally, we choose the state estimate associated with the choice of (ξ, γ) that minimizes $MSE_I(\xi, \gamma)$ over $\mathfrak{H}^{hyper}$.

Recalling Proposition 3.6, if $\{x_i\}_i$ are drawn from an uniform distribution over Ω and the disturbances are homoscedastic, this choice of the hyper-parameters asymptotically minimizes the L^2 state-estimation error. This result holds independently of the strategy employed to compute the state estimate and thus independently of the strategy employed to select the training observation centers. Motivated by this observation, in this work, we choose an uniform grid of sensors for training and we choose the validation sensors by sampling uniformly over Ω. As discussed in [29], if $u^\star_{\xi,\gamma}$ is an accurate description of the true field u^{true}, MSE_I rapidly converges to its expected value. Therefore, the number I of measurements that should be reserved for validation is modest.

Cross-validation is based on the partition of the dataset $\mathcal{D}_L$ into κ equal-sized subsamples (folds) $\{\mathcal{D}_L^{(k)}\}_{k=1}^\kappa$. Of the κ folds, a single fold is retained for testing and the remaining $\kappa - 1$ folds are used for training. The procedure is then repeated κ times with each of the κ folds used once as the validation dataset. In the limit $L = \kappa$, the procedure is known as Leave-One-Out Cross-Validation (LOOCV).

We comment on the application of κ-fold Cross-Validation to the APBDW framework. First, we observe that, even for moderate L, κ-fold Cross-Validation can be quite expensive if $\kappa \approx L$. For this reason, generalized cross-validation strategies, which focus on computing computationally inexpensive approximations of the error indicator, have been developed. We refer to [15, Chapter 7.10] and to the references therein for further details. Second, since in our setting sensor locations are not chosen randomly, it seems very hard to justify cross-validation through a rigorous mathematical argument.

In this paper, we exclusively employ holdout validation and we refer to a future work for the application of more advanced cross-validation strategies.

5. Numerical results for two synthetic problems

In this section, we illustrate the behavior of the APBDW formulation through two two-dimensional model problems: an acoustic Helmholtz problem[1], and a heat transfer problem[2].

Before presenting the results, we briefly summarize some details concerning the implementation of our method. In both tests, we employ csRBF with $k = 1$; recalling Proposition 2.9, this corresponds to $\mathcal{U} = H^{2.5}(\mathbb{R}^2)$. We rely on holdout validation for the choice of ξ and of the kernel parameter γ: we consider uniform grids of training observation points $\{x_m\}_{m=1}^M$, and uniformly random generated validation points $\{x_i\}_{i=1}^I$. In all our tests, we consider $I = \frac{M}{2}$. Performances are measured by computing the mean L^2 error (35).

[1] This model problem is the same considered in [24, Section 3].

[2] This model problem is a slight modification of thermal block problem studied in [26].

5.1. **Application to a synthetic acoustic problem**

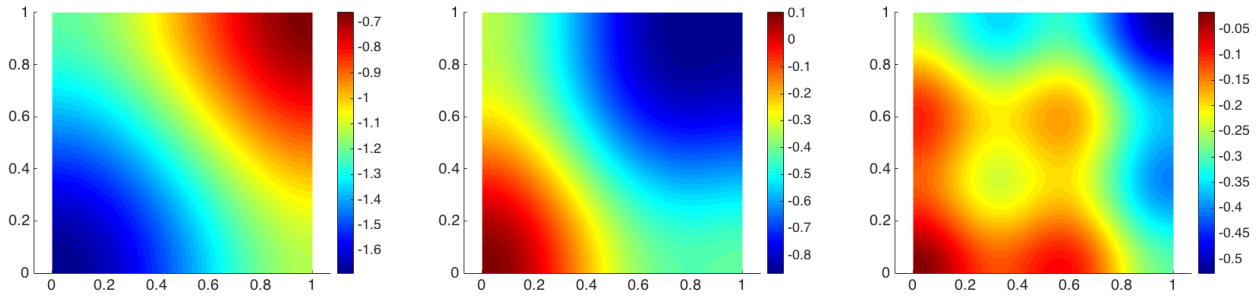
5.1.1. *Problem definition*

Given the domain $\Omega = (0,1)^2$, we define the acoustic model problem:

$$\begin{cases} -(1+i\epsilon\mu)\,\Delta u(\mu) \,-\, \mu^2 u(\mu) = \,\mu\left(2x_1^2 + e^{x_2}\right) & \text{in } \Omega, \\ \partial_n\, u(\mu) = 0 & \text{on } \partial\Omega, \end{cases} \tag{39}$$

where $\mu > 0$ is the wave number and $\epsilon = 10^{-2}$ is a fixed dissipation.

To assess the performance of the APBDW formulation for various configurations, we define the true field u^{true} as the solution to (39) for some $\mu^{true} \in \mathcal{P} = [2, 10]$. Figure 1 shows the true field for three choices of the wave number μ. We approximate the solution using a triangular $\mathbb{P}^5$ finite element discretization.



(A) $Real(u^{true})$, $\mu = 2$ (B) $Real(u^{true})$, $\mu = 5$ (C) $Real(u^{true})$, $\mu = 10$

FIGURE 1. Acoustic synthetic problem: visualization of the truth solutions associated with the synthetic Helmholtz problem.

We introduce some definitions. We consider noisy observations with additive Gaussian noise:

$$y_\ell = u^{true}(x_\ell) + \epsilon_\ell, \qquad \epsilon_\ell \overset{iid}{\sim} \mathcal{N}(0, \sigma^2). \tag{40}$$

Then, we introduce the background spaces $\{\mathcal{Z}_N\}_N$, built using the weak-Greedy algorithm. As mentioned before, we employ csRBF with $k = 1$; recalling Proposition 2.9, this corresponds to $\mathcal{U} = H^{2.5}(\mathbb{R}^2)$.

5.1.2. *Results*

Figure 2(a) investigates the convergence with N for fixed number of sensors M in the absence of noise ($\sigma = 0$). To assess the performance, we compute the relative L^2 error averaged over 20 fields associated with different choices of the parameter μ:

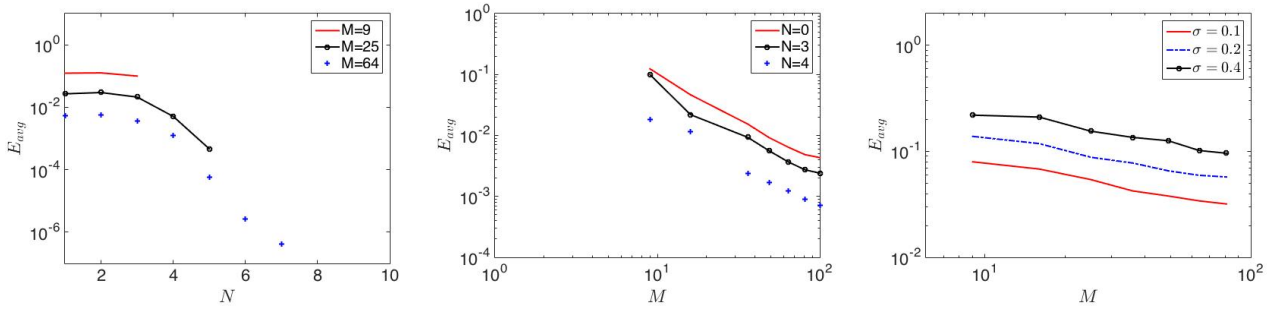
$$E_{avg} := \frac{1}{|\mathcal{P}_{train}|} \sum_{\mu \in \mathcal{P}_{train}} \frac{\|u^{true}(\mu) - u^\star_\xi(\mu)\|_{L^2(\Omega)}}{\|u^{true}(\mu)\|_{L^2(\Omega)}} \tag{41}$$

We observe monotone convergence with respect to N of E_{avg} in the absence of noise.

Figure 6(b) shows the convergence with M for fixed N and noise-free measurements. We assess performances by computing E_{avg} in (41) averaged over 20 fields. We observe that rate of convergence with M weakly depends on the value of N: in this test, we observe $E_{avg} \simeq M^{-1.4}$ for $N = 0$, $E_{avg} \simeq M^{-1.5}$ for $N = 3$, and $E_{avg} \simeq M^{-1.4}$ for $N = 4$. This confirms the multiplicative effect between N convergence and M convergence

observed in Remark 3.4. We further observe that convergence with M is higher than the expected $M^{-1.25}$ (see (30), $\tau = 2.5$). This might be related to the validation procedure for γ that effectively makes the inner product depend on M.

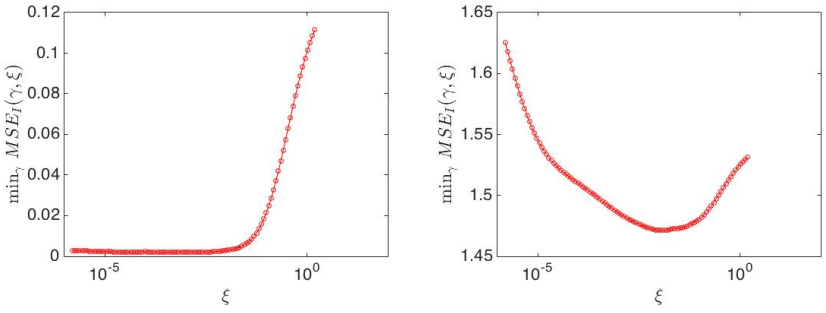
In Figure 6(c), we study the convergence with M in the noisy case. As in the previous tests, we assess performances by computing E_{avg} in (41) averaged over 10 different fields associated with choices of the parameter μ and over 15 realizations of the random noise. We consider the background $\mathcal{Z}_{N=2}$. We observe that the estimated convergence rate in the noisy case is $M^{-0.45}$ for all values of standard deviations σ considered. Interestingly, we observe that the convergence rate does not depend on σ.



(A) (B) (C)

FIGURE 2. Acoustic synthetic problem. Figure (a): convergence with N for fixed M. Figure (b): convergence with M for fixed N. Figure (c): convergence with M for noisy data ($N = 2$).

We finally investigate the connection between the optimal value of ξ and the signal-to-noise ratio. In Figure 3, we compute the mean squared error over the validation set for the estimation of the state associated with the parameter $\mu = 3.68$. We consider $M = 64$ and we compute the mean squared error based on $I = 32$. For this test, we employ the background $\mathcal{Z}_{N=2}$. While for $\sigma = 0.05$ the mean squared error is stable as $\xi \to 0^+$, for $\sigma = 1$ the mean squared error increases as $\xi \to 0^+$ and reaches a minimum for $\xi \approx 10^{-2}$. Since the level of noise is in practice unknown, this numerical result empirically motivates the importance of adapting the value of ξ.



(A) $\sigma = 0.05$ (B) $\sigma = 1$

FIGURE 3. Acoustic synthetic problem: behavior of $\min_\gamma MSE(\xi, \gamma)$ for two different noise-levels.

5.2. **Application to a synthetic non-smooth problem**

5.2.1. *Problem definition*

In this section, we consider the problem:

$$\begin{cases} -\nabla \cdot (\kappa(\mu)\nabla u(\mu)) = 0 & \text{in } \Omega \\ \kappa(\mu)\frac{\partial u}{\partial n} = g & \text{on } \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \\ u(\mu) = 0 & \text{on } \Gamma_4 \end{cases} \tag{42a}$$

where

$$\kappa(x,\mu) = \begin{cases} 1 & \text{in } \Omega_1, \\ \mu_i & \text{in } \Omega_{i+1},\ i = 1,\dots,8; \end{cases} \qquad g(x) = \begin{cases} 1 & \text{on } \Gamma_1, \\ 0 & \text{on } \Gamma_2, \\ 1 - 2x_1 & \text{on } \Gamma_3. \end{cases} \tag{42b}$$

The subdomains $\{\Omega_i\}_i$ and the edges $\{\Gamma_j\}_j$ are shown in Figure 4, while the parameter μ belongs to the compact set $\mathcal{D} = [0.5, 2.5]^8$. Figure 5 shows the true field for three different choices of the parameters[3]. Computations are based on a continuous $\mathbb{P}^3$ Finite Element discretization.

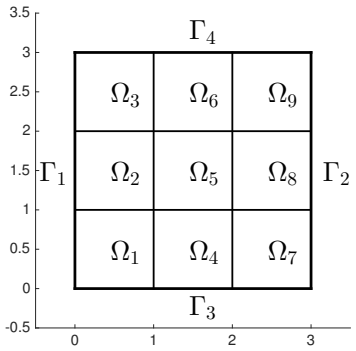


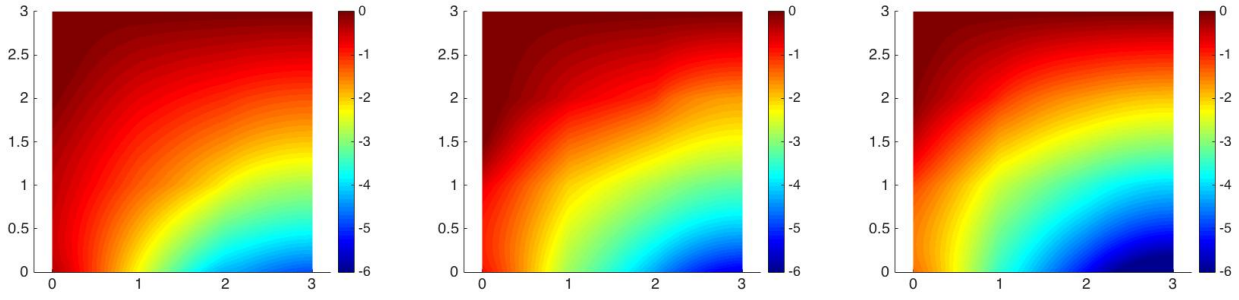
FIGURE 4. Thermal block synthetic problem: computational domain

Unlike the acoustic problem considered before, here the solution does not belong to the native space $\mathcal{U} = H^{2.5}(\mathbb{R}^2)$ associated with the kernel employed. By applying APBDW to this problem, we want to assess how our method performs when applied to non-smooth fields. As in the previous example, we consider noisy observations with additive Gaussian noise (40). Then, we introduce the background spaces $\{\mathcal{Z}_N\}_N$, built using the weak-Greedy algorithm.

5.2.2. *Results*

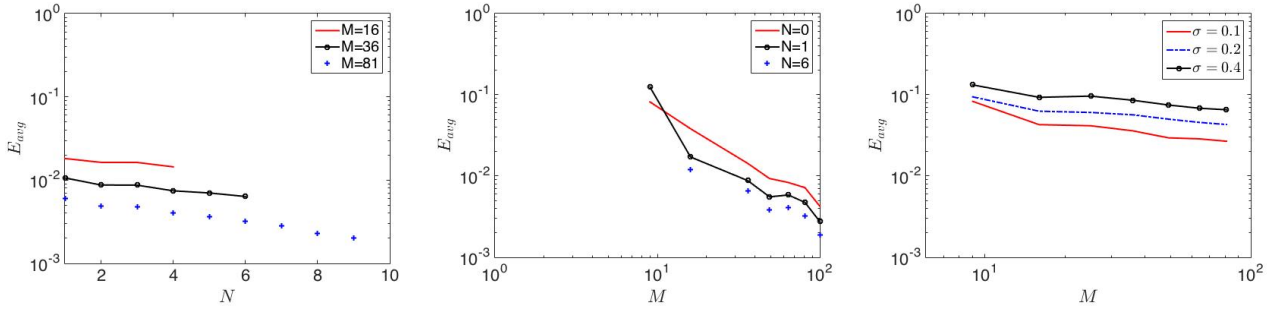
Figure 6 replicates the results shown in Figure 2 for the thermal block problem. As in the previous case, in Figure 6(a), we observe monotone convergence with respect to N of E_{avg} in the absence of noise. In Figure 6(b), we observe $E^2_{avg} \simeq M^{-1.00}$ for $N = 0, 1, 6$. As expected, due to the lack of regularity of the state field, convergence rate is significantly lower compared to the previous example. However, we still observe a multiplicative effect between N convergence and M convergence. Finally, in Figure 6(c), we observe that the estimated convergence rate in the noisy case is $M^{-0.3}$ for all values of the standard deviations considered ($\sigma = 0.1, 0.2, 0.4$). In the noisy case, convergence rate is comparable with the previous example.

[3]The values of the parameter are $\mu^1 = [1.4418, 1.3487, 1.0296, 1.9081, 1.8933, 1.7704, 1.3921, 1.1642]$, $\mu^2 = [0.5672, 2.0695, 1.7911, 1.3486, 2.4778, 1.4982, 1.1617, 0.5927]$, $\mu^3 = [0.6360, 1.4003, 1.5401, 1.1370, 0.9188, 1.2972, 0.9252, 0.7558]$.



(A) $u(\mu^1)$ (B) $u(\mu^2)$ (C) $u(\mu^3)$

FIGURE 5. Thermal block synthetic problem: example of the true field for three values of the parameters.



(A) (B) (C)

FIGURE 6. Thermal block synthetic problem. Figure (a): convergence with N for fixed M. Figure (b): convergence with M for fixed N. Figure (c): convergence with M for noisy data ($N = 2$).

6. Application to a thermal patch configuration

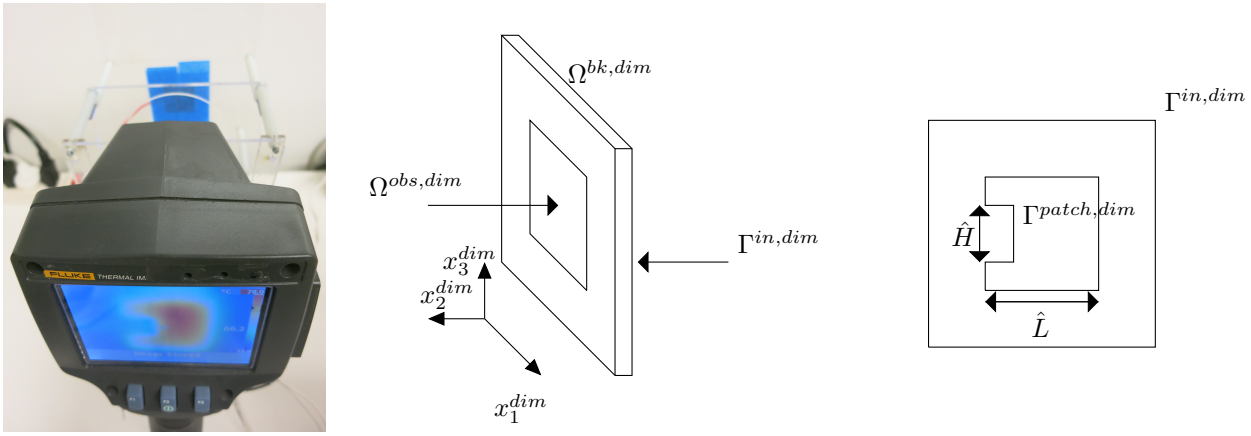
6.1. **Description of the experiment**

The thermal patch system consists of a $1.5mm$ thick acrylic sheet heated from behind by a resistive patch. Heat is generated through an electrical resistance with input power equal to $0.667W$. The goal of the data assimilation procedure is to estimate the temperature field over a portion $\Omega^{obs,dim}$ of the external surface of the plate at the steady-state limit.

We now describe the data acquisition procedure. We use an IR camera (`Fluke Ti 9`) to take measurements in the rectangular region $\Omega^{obs,dim} = [-23.85, 23.85] \times [-17.85, 17.85](mm)$ centered on the patch. Figure 7(a) shows the IR camera. After the patch power is turned on, we take measurements using a sampling time of 4 seconds until steady state is reached; the total duration of the experiment is roughly 5 minutes. The external temperature is about 20^oC, roughly constant throughout the experiment. Each surface measurement taken from the IR camera corresponds to 160×120 pixel-wise measurements; the pixel size is roughly $\Delta h^{device} = 0.3mm$, which is much smaller than the spatial length scale of the phenomenon of interest.

In view of the mathematical description of the problem, we present formal definitions for the geometric quantities involved. First, we introduce the domain $\Omega^{bk,dim} \subset \mathbb{R}^3$ corresponding to the three-dimensional

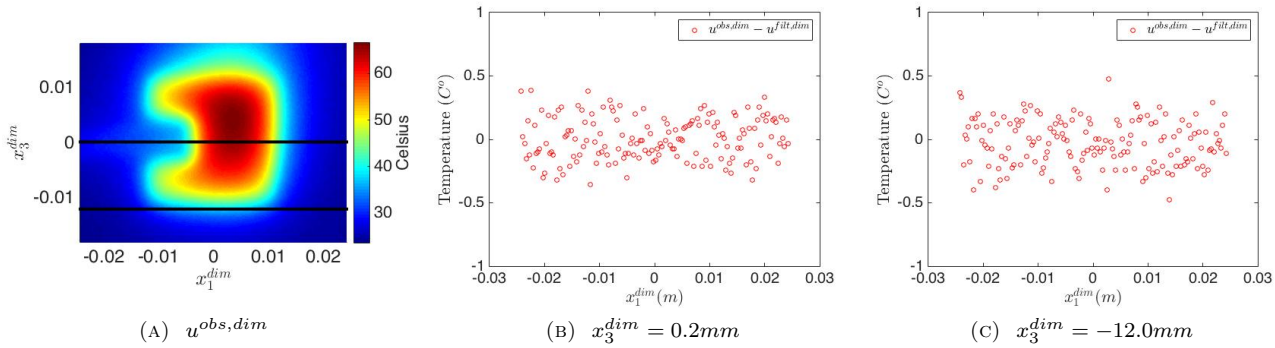
acrylic sheet. We denote by $\Gamma^{patch,dim} \subset \mathbb{R}^2$ the surface of the sheet attached to the patch, and we denote by $\Gamma^{in,dim}$ the face of the sheet that contains $\Gamma^{patch,dim}$. We recall that $\Omega^{obs,dim} \subset \partial\Omega^{bk,dim}$ is the region in which the IR camera takes measurements. Then, we introduce the Cartesian coordinate system $x_1^{dim} x_2^{dim} x_3^{dim}$; according to our definitions, the IR camera takes measurements in the $x_1^{dim} x_3^{dim}$ plane. Figures 7(b) and (c) clarify the definitions of $\Omega^{obs,dim}$, $\Omega^{bk,dim}$, $\Gamma^{patch,dim}$ and $\Gamma^{in,dim}$ and show the characteristic dimensions of the patch.



(A) (B) (C)

FIGURE 7. Thermal patch problem. Figure (a): IR camera. Figures (b) and (c): mathematical description of the acrylic sheet. $\hat{L} = 22.606mm$, $\hat{H} = 9.271mm$.

We now briefly comment on the noise in the dataset and we define the true field. In Figure 8, we show two spatial slices of the field $u^{obs,dim} - u^{filt,dim}$. The field $u^{obs,dim}$ is obtained directly from the IR camera, while $u^{filt,dim}$ is obtained applying a Wiener filter (see, e.g., [20]) based on a 3 by 3 pixel averaging to the field $u^{obs,dim}$. Comparing $u^{filt,dim}$ and $u^{obs,dim}$, we can deduce that the magnitude of noise in the measurements is approximately $\pm 0.5^o C$, roughly independent of the spatial position.



(A) $u^{obs,dim}$ (B) $x_3^{dim} = 0.2mm$ (C) $x_3^{dim} = -12.0mm$

FIGURE 8. Thermal patch problem: comparison between filtered and unfiltered fields. Figure (a): observed thermal field $u^{obs,dim}$. Figures (b) and (c): spatial slices of the difference $u^{obs,dim} - u^{filt,dim}$.

6.2. **Engineering motivation**

We shall now motivate this model problem from the engineering standpoint. Full-field information is typically not available; in practical applications, we envision a system with a local sensor or a small sensor array. For this reason, we want to design a data assimilation state estimation procedure that is able to reconstruct the full field based on a small amount of local measurements.

Since the IR camera provides full-field information, in this work, we synthetize local measurements – the experimental input to our methods – from the IR camera to obtain $\ell_m^{obs} = \ell(u^{obs}, x_m^{obs})$ where the observation functional $\ell(\cdot, x_m^{obs})$ is designed to represent our *fictitious* measurement in the sensor location $x_m^{obs} \in \Omega^{obs}$. We observe that the IR camera permits us to conduct convergence studies that would typically not be feasible in actual field deployment.

6.3. **Mathematical model and background space**

We first present the parametrized best-knowledge model in dimensional form: find $u^{bk,dim} : \Omega^{bk,dim} \mapsto \mathbb{R}$ such that

$$\begin{cases} -\Delta u^{bk,dim} = 0, & \text{in } \Omega^{bk,dim}, \\ \kappa \partial_n u^{bk,dim} + \gamma(u^{bk,dim} - \Theta^{room,dim}) = g^{dim} \chi_{\Gamma^{patch,dim}} & \text{on } \Gamma^{in,dim}, \\ \kappa \partial_n u^{bk,dim} = 0 & \text{on } \partial\Omega^{bk,dim} \setminus \Gamma^{in,dim}, \end{cases} \tag{43}$$

where γ is the convective heat transfer coefficient, κ is the thermal conductivity, $\Theta^{room,dim} = 20^o C$ is the room temperature, and g^{dim} is the incoming flux modeling the heat exchange between the patch and the plate. Textbook values for the model parameters are $\kappa = 0.2 W/m$, $\gamma = 10 W/m^2$.

We now adimensionalize the equation and we define the best-knowledge manifold. Towards this end, we define

$$u^{bk}(x) = \frac{u^{bk,dim}(\hat{L}x) - \Theta^{room,dim}}{\Delta\Theta}, \tag{44}$$

where $\Delta\Theta = 50^o C$ is a rough approximation of the temperature difference between the far-field and the center of the patch, $\hat{L} = 22.606 mm$ is the length of the edge of the patch (see Figure 7). We observe that $u^{bk} = u^{bk}(\mu)$ satisfies

$$\begin{cases} -\Delta u^{bk}(\mu) = 0, & \text{in } \Omega^{bk}, \\ \partial_n u^{bk}(\mu) + \mu u^{bk}(\mu) = g & \text{on } \Gamma^{in}, \\ \partial_n u^{bk} = 0 & \text{on } \partial\Omega^{bk} \setminus \Gamma^{in}, \end{cases} \tag{45a}$$

where $\mu = \hat{L}\gamma/\kappa \approx 1.13$ and g is defined as follows:

$$g(x) = C\, \chi_{\Gamma^{patch}}(x). \tag{45b}$$

Since the model is linear with respect to C and our ultimate goal is to define a linear space associated with the best-knowledge manifold, we can simply set $C = 1$. Assuming that the estimate of κ is accurate and that $\gamma \approx 10 \pm 5 W/m^2$, we have that $\mu \in \mathcal{P} = [0.5650, 1.650]$. We can thus define the best-knowledge manifold as follows:

$$\mathcal{M}^{bk} = \left\{ u^{bk}(\mu)|_{\Omega^{obs}} : \mu \in \mathcal{P} \right\}. \tag{46}$$

Some comments are in order. Parametric uncertainty in the model is associated with the value of μ (that is, with the convective heat transfer coefficient and the thermal conductivity), while non-parametric uncertainty is mainly related to the nonlinear effects associated to natural convection and to the heat exchange between the patch and the sheet. To compute the solution to the best-knowledge model, we recur to a $\mathbb{P}^3$ continuous Finite Element discretization based on $\mathcal{N} = 40000$ degrees of freedom.

The background space $\mathcal{Z}_N$ associated with (45)-(46) is built using the weak-Greedy algorithm. Then, we restrict the space $\mathcal{Z}_N^{bk}$ to the domain of interest $\Omega = \Omega^{obs}$ to form $\mathcal{Z}_N$.

6.4. **Numerical results**

Figure 9 shows the convergence with M for the RB space $\mathcal{Z}_N$ for $N = 0$, $N = 1$ and $N = 2$. For this test, we consider holdout validation for ξ and γ with $I = M/2$. On the y-axis, we report the mean squared error computed based on the entire full-field information.

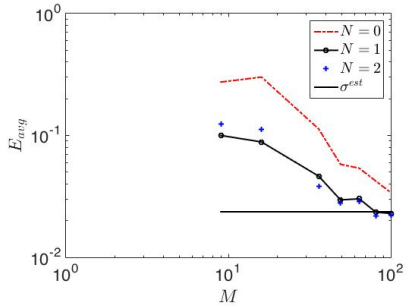


FIGURE 9. Thermal patch problem: convergence with M for fixed N.

We observe that for $M \approx 100$ we reach the noise level

$$\sigma^{est} := \frac{\|\frac{0.5^o C}{\Delta\Theta}\|_{L^2(\Omega)}}{\|u^{obs}\|_{L^2(\Omega)}}.$$

We also observe that, while including the first snapshot leads to a substantial improvement in the performances, considering $N > 1$ does not lead to any substantial improvement.

7. CONCLUSIONS

In this paper, we presented the APBDW approach to the variational data assimilation (state estimation) problem. The approach generalizes the PBDW formulation presented in [24] to the case of pointwise noisy measurements. We also discussed a well-posedness analysis and *a priori* and *a posteriori* error estimates for the L^2 state-estimation error. Our well-posedness analysis also holds when the background space does not belong to the native space induced by the kernel employed. On the other hand, the error analysis relies on the assumption that both the true field and the background belong to the native space. We finally presented synthetic and experimental numerical results to prove the effectiveness of our approach.

We now identify three extensions to the approach, which are subjects of future work. First, we wish to design strategies for the selection of the observation centers that address both stability and approximation. This would tighten the connection between our approach and the design of the experiment. Second, we wish to consider more general observation functionals of the form $y_m = \ell_m(u^{true}) + \epsilon_m$. This would allow us to take into account different sources of information. In this respect, we observe that the well-posedness analysis does not depend on the fact that $y_m = u^{true}(x_m) + \epsilon_m$, while the error analysis relies on the form of the observation functionals. Finally, we wish to exploit the interpretation of APBDW as convex relaxation of the partial spline model for a particular choice of the background to derive new probabilistic and deterministic error bounds and possibly improve the performances of the state-estimation procedure.

Acknowledgments: the author thanks Prof. Anthony T. Patera (MIT) for fruitful discussions, Dr. James D. Penn (MIT) for the experimental results in section 6, and Prof. Masayuki Yano (University of Toronto) for the proof of Lemma 2.2 and for the development of the high-order FE code used in the numerical sections. This

work was supported by OSD/AFOSR/MURI Grant FA9550-09-1-0613, ONR Grant N00014-11-1-0713, and the MIT-Singapore International Design Center.

Appendix A. Proofs of the error bounds

In this appendix, we provide the proofs of the results presented in section 3. The appendix is organized as follows. In section A.1, we introduce a regularized formulation and we show the connection with our formulation. In section A.2, we exploit the results for the regularized problem to prove Proposition 3.1. Finally, in section A.3, we prove the result for fixed-design regression.

A.1. **Preliminaries**

We introduce a regularized formulation of the APBDW statement proposed in this work: given $\lambda \geq 0$, $\xi > 0$, find $u^\star_{\lambda,\xi} \in \mathcal{U}$ such that

$$u^\star_{\lambda,\xi} = \arg\min_{u\in\mathcal{U}} \; J_{\lambda,\xi}(u) := \xi\|u\|^2_{\lambda,N} + V_M(u), \tag{47}$$

where the seminorm $\|\cdot\|_{\lambda,N}$ is defined as

$$\|w\|^2_{\lambda,N} = \lambda\|\Pi_{\mathcal{Z}_N} w\|^2 + \|\Pi_{\mathcal{Z}_N^\perp} w\|^2. \tag{48}$$

We observe that for any $\lambda > 0$, the function $\|\cdot\|_{\lambda,N}$ is a norm equivalent to $\|\cdot\|$. We also observe that for $\lambda = 0$, problem (47) corresponds to (2).

Next proposition summarizes a number of properties of problem (47) that are crucial to prove the error bounds for $u^\star_\xi$.

Proposition A.1. *Let $\beta_{N,M} > 0$. Then, the following hold.*

1. *For any $\lambda > 0$, the solution to (47) exists and is unique. Furthermore, if we introduce $\eta^\star_{\lambda,\xi} = \Pi_{\mathcal{Z}_N^\perp} u^\star_{\lambda,\xi}$, $z^\star_{\lambda,\xi} = \Pi_{\mathcal{Z}_N} u^\star_{\lambda,\xi}$, we have that $\eta^\star_{\lambda,\xi} \in$ span$\{\Pi_{\mathcal{Z}_N^\perp} K_{x_m}\}_{m=1}^M$ and $z^\star_{\lambda,\xi} \in$ span$\{\Pi_{\mathcal{Z}_N} K_{x_m}\}_{m=1}^M$.*
2. *For any $\xi > 0$, the solution $u^\star_{\lambda,\xi}$ converges to the solution $u^\star_\xi$ to (2) when $\lambda \to 0^+$.*
3. *For any $\lambda \geq 0$, the following bounds hold:*

$$\|u^{true} - u^\star_{\lambda,\xi}\|_{\lambda,N} \leq 2\|u^{true}\|_{\lambda,N} + \frac{\delta}{2\sqrt{\xi}}, \tag{49a}$$

and

$$\|u^{true} - u^\star_{\lambda,\xi}\|_{\ell^2(\mathcal{X}_M)} \leq \sqrt{M}\left(\delta + \frac{\sqrt{\xi}}{2}\|u^{true}\|_{\lambda,N}\right). \tag{49b}$$

We prove each statement separately.

Proof. (*statement 1*) For any $\lambda > 0$, $u \mapsto \|u\|^2_{\lambda,N}$ is strictly convex, while $u \mapsto V_M(u)$ is convex. This implies that for any $\xi > 0$ the objective function $J_{\lambda,\xi}(u) = \xi\|u\|^2_{\lambda,N} + V_M(u)$ is strictly convex. Therefore, existence and uniqueness of the solution to (47) follow from [10, Theorem 3, Chapter 8.2].

We observe that $x \in \Omega \mapsto \Phi^{\lambda,N}_x = \frac{1}{\lambda}\Pi_{\mathcal{Z}_N} K_x + \Pi_{\mathcal{Z}_N^\perp} K_x$, is the feature map associated with $(\mathcal{U}, \|\cdot\|_{\lambda,N})$. We have indeed that for all $v \in \mathcal{U}$ and $x \in \Omega$

$$(\Phi^{\lambda,N}_x, v)_{\lambda,N} = \frac{\lambda}{\lambda}(\Pi_{\mathcal{Z}_N} K_x, v) + (\Pi_{\mathcal{Z}_N^\perp} K_x, v) = (K_x, \Pi_{\mathcal{Z}_N} v + \Pi_{\mathcal{Z}_N^\perp} v) = (K_x, v) = v(x).$$

Exploiting the representer theorem (see, e.g., [34, Theorem 16.1]), we have that $u^\star_{\lambda,\xi} \in \text{span}\{\Phi^{\lambda,N}_{x_m}\}_{m=1}^M$. As a result, we have that $\eta^\star_{\lambda,\xi} \in \text{span}\{\Pi_{\mathcal{Z}_N^\perp} \Phi^{\lambda,N}_{x_m}\}_{m=1}^M$, and $z^\star_{\lambda,\xi} \in \text{span}\{\Pi_{\mathcal{Z}_N} \Phi^{\lambda,N}_{x_m}\}_{m=1}^M$ for any $\lambda > 0$. □

Proof. (*statement 2*) Let $\{\lambda_j\}_j$ be a real sequence such that $\lambda_j \to 0^+$. Exploiting the first statement of Proposition A.1, we have that sequences $\{\eta^\star_{\lambda_j,\xi}\}_j$, $\{z^\star_{\lambda_j,\xi}\}_j$ belong to finite dimensional spaces that do not depend on λ. Furthermore, applying Lemma 2.3, we have that they are uniformly bounded for all j. Applying Bolzano-Weierstrass theorem, the sequence $\{u^\star_{\lambda_j,\xi} = \eta^\star_{\lambda_j,\xi} + z^\star_{\lambda_j,\xi}\}_j$ admits a strongly convergent subsequence $\{u^\star_{\lambda_k,\xi}\}_k$ to $\widehat{u}^\star_\xi \in \mathcal{U}$.

We now show that $\widehat{u}^\star_\xi = u^\star_\xi$. We first observe that

$$J_{\lambda_k,\xi}(u^\star_{\lambda_k,\xi}) = \lambda_k \underbrace{\|z^\star_{\lambda_k,\xi}\|^2}_{\leq C} + \|\eta^\star_{\lambda_k,\xi}\|^2 + V_M(u^\star_{\lambda_k,\xi}) \to J_\xi(\widehat{u}^\star_\xi), \qquad k \to \infty.$$

We further observe that for any $\lambda_k > 0$

$$J_{\lambda_k,\xi}(u^\star_{\lambda_k,\xi}) \leq J_{\lambda_k,\xi}(u^\star_\xi), \qquad k = 1, 2, \dots,$$

and by taking the limit on both sides, we obtain

$$J_\xi(\widehat{u}^\star_\xi) \leq J_\xi(u^\star_\xi)$$

Since $u^\star_\xi$ is the unique minimizer of (2), we must have $u^\star_\xi = \widehat{u}^\star_\xi$. Furthermore, by the same argument, $u^\star_\xi$ must be the only limit point of the sequence; therefore, the entire sequence converges to $\widehat{u}^\star_\xi$. Thesis follows. □

Proof. (*statement 3*) For $\lambda > 0$, $\|\cdot\|_{\lambda,N}$ is a norm for $\mathcal{U}$; therefore, estimates (49a) and (49b) follow directly from [18, Corollary 4.3] and [18, Lemma 4.5].

The extension to $\lambda = 0$ follows by observing that $u^\star_{\lambda,\xi}$ converges to $u^\star_\xi$ when $\lambda \to 0^+$. □

Before proving the error bounds, we prove (26b).

Lemma A.2. *Let Ω be a Lipschitz domain and let $\mathcal{U}$ be the Sobolev space $H^\tau(\Omega)$ with $\tau > d/2$. Let us assume that inf-sup constant $\beta_{N,M}$ defined in (3) is strictly positive and $h_{\mathcal{X}_M} < 1$.*

Then,

$$C_{N,\mathcal{X}_M} \leq \frac{1}{\min\{c_{N,M}, 1 - h^{2\tau-d}_{\mathcal{X}_M}\}}\, C,$$

where $c_{N,M}$ is defined in (6) and C depends on the domain Ω and on $(\cdot,\cdot)$.

Proof. Let us define the constant

$$\widehat{C}_{\mathcal{X}_M} := \sup_{u \in \mathcal{U}} \frac{\|u\|^2_{L^2(\Omega)}}{h^{2\tau}_{\mathcal{X}_M}\|u\|^2 + h^d_{\mathcal{X}_M}\|u\|^2_{\ell^2(\mathcal{X}_M)}}.$$

Recalling [18, Theorem 4.8], $\widehat{C}_{\mathcal{X}_M}$ is bounded from above by a constant C that does not depend on M.

Since $\beta_{N,M} > 0$, recalling Lemma 2.3, we have that

$$\|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + \|u\|^2_{\ell^2(\mathcal{X}_M)} \geq c_{N,M}\|u\|^2,$$

where $c_{N,M} > 0$ is given by the expression in (6). Then, we observe that

$$\begin{aligned} h^{2\tau}_{\mathcal{X}_M}\|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + h^d_{\mathcal{X}_M}\|u\|^2_{\ell^2(\mathcal{X}_M)} = &\quad h^{2\tau}_{\mathcal{X}_M}\left(\|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + \|u\|^2_{\ell^2(\mathcal{X}_M)}\right) + \left(h^d_{\mathcal{X}_M} - h^{2\tau}_{\mathcal{X}_M}\right)\|u\|^2_{\ell^2(\mathcal{X}_M)} \\ \geq &\quad c_{N,M} h^{2\tau}_{\mathcal{X}_M}\|u\|^2 + \left(1 - h^{2\tau-d}_{\mathcal{X}_M}\right) h^d_{\mathcal{X}_M}\|u\|^2_{\ell^2(\mathcal{X}_M)} \\ \geq &\quad \min\{c_{N,M}, 1 - h^{2\tau-d}_{\mathcal{X}_M}\}\left(h^{2\tau}_{\mathcal{X}_M}\|u\|^2 + h^d_{\mathcal{X}_M}\|u\|^2_{\ell^2(\mathcal{X}_M)}\right). \end{aligned}$$

As a result,

$$C_{N,\mathcal{X}_M} = \sup_{u\in\mathcal{U}} \frac{\|u\|^2_{L^2(\Omega)}}{h^{2\tau}_{\mathcal{X}_M}\|\Pi_{\mathcal{Z}_N^\perp}u\|^2 + h^d_{\mathcal{X}_M}\|u\|^2_{\ell^2(\mathcal{X}_M)}} \leq \underbrace{\left(\sup_{u\in\mathcal{U}} \frac{\|u\|^2_{L^2(\Omega)}}{h^{2\tau}_{\mathcal{X}_M}\|u\|^2 + h^d_{\mathcal{X}_M}\|u\|^2_{\ell^2(\mathcal{X}_M)}}\right)}_{=C} \frac{1}{\min\{c_{N,M}, 1-h^{2\tau-d}_{\mathcal{X}_M}\}}$$

Thesis follows. □

A.2. **Proof of the error bound for scattered data**

Proof. (*Proposition 3.1*) The proof replicates the argument of [18, Theorem 4.11]. Recalling the definition of $C_{N,\mathcal{X}_M}$, we have

$$\|u^{true} - u^\star_\xi\|^2_{L^2(\Omega)} \leq C_{N,\mathcal{X}_M}\left(h^{2\tau}_{\mathcal{X}_M}\,\|\Pi_{\mathcal{Z}_N^\perp}(u^{true}-u^\star_\xi)\|^2 \,+\, h^d_{\mathcal{X}_M}\,\|u^{true}-u^\star_\xi\|^2_{\ell^2(\mathcal{X}_M)}\right).$$

Then, using (49a) and (49b), we obtain

$$\|u^{true} - u^\star_\xi\|^2_{L^2(\Omega)} \leq C_{N,\mathcal{X}_M}\left(h^{2\tau}_{\mathcal{X}_M}\left(2\,\|\Pi_{\mathcal{Z}_N^\perp}u^{true}\|^2 \,+\, \frac{\delta}{2}\frac{1}{\sqrt{\xi}}\right)^2 + h^d_{\mathcal{X}_M} M\left(\delta + \frac{\sqrt{\xi}}{2}\|\Pi_{\mathcal{Z}_N^\perp}u^{true}\|\right)^2\right),$$

which is the thesis. □

A.3. **Proof of the error bound for fixed-design regression**

We first introduce some notation and preliminary definitions. We decompose the datum $\mathbf{y}$ as

$$\mathbf{y} = \mathbf{y}^{true} + \boldsymbol{\epsilon}, \qquad \mathbf{y}^{true} = [u^{true}(x_1),\dots,u^{true}(x_M)],\;\; \boldsymbol{\epsilon} = [\epsilon_1,\dots,\epsilon_M],$$

and we define $\boldsymbol{\epsilon}_{aug} = \begin{bmatrix} \boldsymbol{\epsilon} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{M+N}$. We observe that $Var(\boldsymbol{\epsilon}_{aug}) = \sigma^2\Sigma$, where Σ is defined in(31). Then, we introduce the solution $u^{\star,\sigma=0}_\xi$ to (2) for $\mathbf{y} = \mathbf{y}^{true}$. We further introduce the vectors of coefficients $\mathbf{u}^\star, \mathbf{u}^{\star,\sigma=0} \in \mathbb{R}^{M+N}$,

$$\mathbf{u}^\star = \begin{bmatrix} \boldsymbol{\eta}^\star \\ \boldsymbol{z}^\star \end{bmatrix}, \quad \mathbf{u}^{\star,\sigma=0} = \begin{bmatrix} \boldsymbol{\eta}^{\star,\sigma=0} \\ \boldsymbol{z}^{\star,\sigma=0} \end{bmatrix},$$

associated with $u^\star_\xi$ and $u^{\star,\sigma=0}_\xi$.

We have now the elements to prove Proposition 3.5.

Proof. (*Proposition 3.5*) We observe that

$$\|u^\star_\xi - u^{\star,\sigma=0}_\xi\|^2_{L^2(\Omega)} = (\mathbf{u}^\star - \mathbf{u}^{\star,\sigma=0})^T\,\mathbb{M}\,(\mathbf{u}^\star - \mathbf{u}^{\star,\sigma=0}) = \boldsymbol{\epsilon}^T_{aug}\left(\mathbb{A}_\xi^{-1}\,\mathbb{M}\,\mathbb{A}_\xi^{-1}\right)\boldsymbol{\epsilon}_{aug}.$$

Then, applying [27, Theorem C, Chapter 14.4], we find

$$\mathbb{E}\left[\|u^\star_\xi - u^{\star,\sigma=0}_\xi\|^2_{L^2(\Omega)}\right] = \sigma^2\,\mathrm{trace}\left(\mathbb{A}_\xi^{-1}\,\mathbb{M}\,\mathbb{A}_\xi^{-1}\,\Sigma\right). \tag{50}$$

Thesis follows by observing that

$$\mathbb{E}\left[\|u^\star_\xi - u^{true}\|^2_{L^2(\Omega)}\right] \leq 2\,\|u^{true} - u^{\star,\sigma=0}_\xi\|^2_{L^2(\Omega)} + 2\,\mathbb{E}\left[\|u^\star_\xi - u^{\star,\sigma=0}_\xi\|^2_{L^2(\Omega)}\right]$$

and then combining estimates (30) and (50). □

References

[1] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, pages 337–404, 1950.

[2] A. Bennett. Array design by inverse methods. *Progress in oceanography*, 15(2):129–156, 1985.

[3] A. Bennett and P. McIntosh. Open ocean modeling as an inverse problem: tidal theory. *Journal of Physical Oceanography*, 12(10):1004–1018, 1982.

[4] A. F. Bennett. *Inverse modeling of the ocean and atmosphere*. Cambridge University Press, 2002.

[5] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Data assimilation in reduced modeling. *arXiv preprint arXiv:1506.04770*, 2015.

[6] Y. Cao, J. Zhu, I. M. Navon, and Z. Luo. A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *International Journal for Numerical Methods in Fluids*, 53(10):1571–1583, 2007.

[7] A. Cohen and R. DeVore. Approximation of high-dimensional parametric pdes. *Acta Numerica*, 24:1–159, 5 2015.

[8] P. Courtier, J.-N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387, 1994.

[9] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bulletin of AMS*, 2001.

[10] L. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 1998.

[11] R. Everson and L. Sirovich. Karhunen–loeve procedure for gappy data. *JOSA A*, 12(8):1657–1664, 1995.

[12] T. Gasser and H. G. Müller. *Kernel estimation of regression functions*. Springer, 1979.

[13] L. Györfi, M. Kohler, A. Krzyzak, and H. Walk. *A distribution-free theory of nonparametric regression*. Springer Science & Business Media, 2006.

[14] E. Hart, S. Cox, K. Djidjeli, and V. Kubytskyi. Solving an eigenvalue problem with a periodic domain using radial basis functions. *Engineering analysis with boundary elements*, 33(2):258–262, 2009.

[15] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning*, volume 2. Springer, 2009.

[16] G. Kimeldorf and G. Wahba. Some results on tchebycheffian spline functions. *Journal of mathematical analysis and applications*, 33(1):82–95, 1971.

[17] R. Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145, 1995.

[18] J. Krebs, A. Louis, and H. Wendland. Sobolev error estimates and a priori parameter selection for semi-discrete tikhonov regularization. *Journal of Inverse and Ill-Posed Problems*, 17(9):845–869, 2009.

[19] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM Journal on Numerical analysis*, 40(2):492–515, 2002.

[20] J. S. Lim. *Two-dimensional signal and image processing*. Prentice Hall, 1990.

[21] A. Lorenc. A global three-dimensional multivariate statistical interpolation scheme. *Monthly Weather Review*, 109(4):701–721, 1981.

[22] A. C. Lorenc. Analysis methods for numerical weather prediction. *Royal Meteorological Society, Quarterly Journal*, 112:1177–1194, 1986.

[23] Y. Maday and O. Mula. A generalized empirical interpolation method: Application of reduced basis techniques to data assimilation. In *Analysis and Numerics of Partial Differential Equations*, pages 221–235. Springer, 2013.

[24] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano. A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics. *International Journal for Numerical Methods in Engineering*, 2014.

[25] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano. Pbdw state estimation: Noisy observations; configuration-adaptive background spaces; physical interpretations. *ESAIM: Proceedings and Surveys*, 50:144–168, 2015.

[26] A. T. Patera and G. Rozza. Reduced basis approximation and a posteriori error estimation for parametrized partial differential equations, 2007.

[27] J. Rice. *Mathematical statistics and data analysis*. Nelson Education, 2006.

[28] G. Rozza, D. P. Huynh, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods in Engineering*, 15(3):229–275, 2008.

[29] T. Taddei, J. D. Penn, and A. T. Patera. Experimental a posteriori error estimation by monte carlo sampling of observation functionals. Technical report, MIT, 2015. submitted to M3AS.

[30] V. Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[31] P. Vermeulen and A. Heemink. Model-reduced variational data assimilation. *Monthly weather review*, 134(10):2888–2899, 2006.

[32] G. Wahba. Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 364–372, 1978.

[33] G. Wahba. *Spline models for observational data*, volume 59. Siam, 1990.

[34] H. Wendland. *Scattered data approximation*, volume 17. Cambridge university press, 2004.

[35] K. Willcox. Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. *Computers & fluids*, 35(2):208–226, 2006.

[36] M. Yano. Private communication. June 2015.